LARGE-SCALE STRUCTURE OF THE UNIVERSE: SIMULATIONS AND STATISTICS

By

MICHAEL A. JONES

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1997

Dedicated to my brother Oggie
George Clinton, Bernie Worrell, Bootsy Collins, and Bill Laswell
and Max the friendliest cat

ACKNOWLEDGMENTS

Most importantly I wish to thank my parents. This is really dedicated to them as they have always been so very supportive of me.

Despite my ascension to the highest Cabinet post, Slackmaster General, my friends and relatives have always been polite and encouraging. My advisor Jim Fry deserves special thanks for being patient and friendly. It is an honor to have such an excellent physicist for an advisor. Professor Neil Sullivan (head of the department), the professors on the graduate committee, and the entire staff at the Physics Department have been very helpful in allowing me to complete my degree. Friends from the Physics Department, Erika Kisvarsanyi, Dave Willmes, and Steve Thomas have provided assistance (steaming the mad cow). Particular thanks are due to Chandra Chegireddy, who is to the Physics Department what Scottie is to the Enterprise.

My good friends Oggie Jones and Bill Hoston share the credit (blame) for influencing me to do physics. Without their good and fun company, I would have bowed out of technical endeavors long ago. As I completed my dissertation, several friends taught me valuable skills: Deborah and Kerry Duggan—how to ride bikes, Karen and Bobby Livingston—how to lift weights, Professor Vincent Rodgers—how to place theoretical physics, computers, bass, and drums into a proper perspective.

This research has made use of NASA's Astrophysics Data System Abstract Service. This document was typeset using plain TEX. Computer simulations were performed on DEC Alpha workstations.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS iii

ABSTRACT vi

CHAPTERS

1 VISUALIZING LARGE-SCALE STRUCTURE AN ELEMENTARY INTRODUCTION 1

1.1 Introduction 1
1.2 Beyond the Night Sky 2
1.3 Galaxies and Our Neighborhood of the Universe 4
1.4 Structures in an Expanding Universe 15
1.5 Mapping the Large-scale Structure 39
1.6 Statistics and Simulations 50
1.7 Postlude 62

2 COSMIC STRUCTURE FORMATION 65

2.1 Big Bang Background 65
2.2 Cosmic Structure 77
2.3 Statistics 101
2.4 Perturbation Theory for Statistics 116

3 ZEL'DOVICH APPROXIMATION 122

3.1 The Equations of Motion and the Approximation 122
3.2 Pancakes and Other Features 133
3.3 Perturbation Theory 148

4 SIMULATIONS 155

4.1 Numerical Realizations 155
4.2 Power and Bias 182

5 STATISTICS 204

5.1 Introduction 204
5.2 Measured Correlations 206
5.3 Nonlinearity and Bias 254

6 FINAL REMARKS 280

REFERENCES 294
BIOGRAPHICAL SKETCH 300

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

LARGE-SCALE STRUCTURE OF THE UNIVERSE
SIMULATIONS AND STATISTICS

By

Michael A. Jones

May 1997

Chairman: J. N. Fry
Major Department: Physics

The gravitational formation of large-scale structure in an expanding universe is studied within the framework of the standard hot big bang model. We work in the scenario whereby small fluctuations in the initial density field are amplified via gravitational instability, generating frothy, sheet-like structures which span scales from a few million to a few hundred million light-years. Perturbation theory is utilized, based on the equations of a collisionless gravitating fluid. Using an efficient approximation which reproduces essential aspects of the complete model, we perform computer simulations of large-scale matter and galaxy distributions. Random realizations of the initial conditions are produced with varying initial statistical properties, and the final, evolved matter distribution is investigated. The power spectrum and bispectrum (the two-point and three-point correlation functions in Fourier space) are measured to statistically characterize the spatial fluctuations. Perturbation theory is tested, and the transition from the linear to nonlinear evolutionary regimes is investigated. It is found that a perturbative approach can be selectively used beyond its expected range of applicability. In some instances, highly structured, yet dilute, galaxy distributions can be usefully distinguished utilizing results from perturbation theory.

CHAPTER 1
VISUALIZING LARGE-SCALE STRUCTURE AN ELEMENTARY INTRODUCTION

1.1 Introduction

Cosmology is the branch of astrophysics concerned with the observable universe in its entirety. The following chapters present computational and theoretical work—as opposed to observational/experimental work—related to the large-scale structure problem of cosmology. The purpose of Chapter 1 is to introduce the concept of large-scale structure in the universe. We wish in Chapter 1 to informally describe the matter distribution on extragalactic scales and the big bang model which provides the backdrop for the clustering of matter. Chapter 2 provides some technical background for our particular area of research: gravitational instability in an expanding universe studied within a Newtonian approximation. While Chapter 1 offers a visual or intuitive picture of cosmic structure, Chapter 2 introduces the correlation statistics and expanding fluid perturbation theory which we use to technically describe the structure and its dynamics. Further details of the perturbative expanding fluid approach are described in Chapter 3, including the simplifying innovation of Zel'dovich (1970). Results from computer simulations are presented in Chapters 4 and 5. Chapter 4 shows examples of computer simulated structure under various conditions, while Chapter 5 provides statistical analyses of simulated cosmic structure. We close with some final remarks in Chapter 6.

1.2 Beyond the Night Sky

At night, we may look out and perceive the vastness of interstellar space. On a clear night, we might see several thousand stars which reside in our neighborhood of the Milky Way, our home galaxy. From an Earthly point of view, these stars are separated by truly enormous distances. Yet from a cosmological standpoint, these interstellar distances are quite small. In order to investigate the nature of the universe on the largest observable scales, one must penetrate far beyond the night sky.

It is fair to say that to us the Sun is very powerful and important. Nonetheless we must accept that it is simply one star among the over one hundred billion stars which inhabit the Milky Way. The galaxy is an interesting and complex system, but we only have to comprehend it in a broad sense before we try to view an even larger, more inclusive, cosmic picture. Our galaxy is just one of billions of galaxies which we can observe from Earth. Each galaxy can be thought of as an "island universe" containing billions of stars. Within the fields of astronomy and astrophysics, the study of large-scale structure is concerned with the distribution and motion of matter on cosmological scales. We are at times interested in galaxy-sized regions. More often than not we seek out the far reaches of intergalactic space, regions containing hundreds, thousands, or even millions of galaxies. Astronomical data show that galaxies are not simply sitting out there in an arbitrary pattern. Essentially all other galaxies appear to be receding from us, a fact which is a natural ingredient in expanding universe (or "big bang") cosmologies. Furthermore, the galaxy positions are clustered in a nontrivial way. What is the nature of galaxy clustering? How are galaxies formed in the first place? These questions lie at the heart of the large-scale structure problem.

There is a hierarchy of structure in the universe. Matter clumps together in stars which in turn reside in galaxies. Then there are galaxy clusters which range from small groups to rich clusters which may contain a thousand galaxies. Even galaxy clusters are found to be grouped together into superclusters which may contain thousands of galaxies. Attention must also be paid to the possibility of dark matter. In the simplest sense, dark matter refers to nonluminous matter, or matter that we cannot see with a telescope.[1] In the same sense that the planet Pluto was discovered before it was seen, there is matter out there on a cosmological scale which we detect (by its gravitational influence) but can not see. Any clumping of matter, luminous or dark, on extra-galactic scales is customarily referred to as "structure" or "large-scale structure."

What sort of physical theories will our discussion of cosmological structure formation rely on? In order to imagine a simple scenario which describes the formation of large-scale structure—and which bears some resemblance to observable reality—we must proceed within the framework of an expanding universe model. The standard big bang model of cosmology is more of a framework or paradigm than a totally specific theory. Nonetheless, it is a broad theory which (a) makes predictions, (b) can be tested against observations, (c) can be bent due to its flexibility, and, importantly, (d) can conceivably be disproven. But right now it is really the only player in the game, and so we have to go with it, and we do so without regret, for its successes are quite remarkable. We can think of the standard big bang model as taking known and trusted theories such as thermodynamics, electromagnetism, and particle physics and placing them in the context of an expanding universe. We have not mentioned gravity yet; we have saved the most important for last. In terms of interparticle interactions (involving, say, protons, neutrons, and electrons) gravity is the weakest known force, but it is a long-range force. The "strong force" (which is responsible for the binding of the atomic nucleus) and the "weak force" (responsible for radioactive decay) are both very short-range subatomic forces. The electromagnetic force is long-range, but its effect is literally neutralized with respect to long-range astronomical application.[2] The gravitational force, whether in the guise of Einstein's general theory of relativity or Newton's theory of gravitation, plays a dominant role in cosmology, since we believe that gravity governs the expansion of the universe and is responsible for the formation and clustering of galaxies.

---

[1] To "see" an object in this context means to detect its radiation in any region of the spectrum, from radio to visible to x-ray, etc.

In the remainder of the chapter, we briefly describe the expanding universe model and the large-scale clustering of matter. We begin with our own Milky Way.

1.3 Galaxies and Our Neighborhood of the Universe

1.3.1 An Overview of Galaxies

The Milky Way galaxy is a collection of over one hundred billion stars bound by gravity into a flattened, rotating disk. The sun lies within the disk, about two thirds of the way out from the galactic center. There are thousands of nearby galaxies which are somewhat similar to the Milky Way and which provide a useful basis for comparison. Figure 1-1 shows negative photographic images of four galaxies which lie far outside the boundaries of our galaxy but which are nearby compared to millions of more distant galaxies. However, like most other galaxies, these are too far away for their billions of constituent stars to be individually resolved. The two spiral galaxies (labeled NGC 4902 and NGC 3200) in Figure 1-1 resemble the Milky Way—if an observer from another galaxy were to photograph our galaxy through a telescope, they would see a well defined spiral galaxy. Of course, if this hypothetical observer happened to photograph our galaxy from a side view, the results would be different. Two such side views are shown in Figure 1-2.

---

[2] Consider the case of the Earth and the Moon. The electromagnetic forces between positively charged protons and negatively charged electrons at the atomic and molecular level are quite influential, but overall the Earth is practically neutral (due to the nullifying effect of nearly equal numbers of protons and electrons on Earth). Since the moon is also electrically neutral, only gravity is left to influence their relative motion. This argument extends to all long-range astronomical interactions.

All six galaxy images shown in Figures 1-1 and 1-2 were obtained by the Automatic Plate Measuring Survey, a modern astronomical survey containing over two million galaxies. Even our small sampling of six pictures shows a bit of the diversity in galaxy types. Most of the galaxies on record fall into two broad categories, spiral and elliptical. For example, in Figure 1-1 the galaxy NGC 3962 is elliptical, and NGC 2811 is lenticular, a subclass of spiral. First, we note that such classifications are not merely idle categorization. The galaxy classification schemes reveal complex (and not yet fully understood) correlations between galaxy type and the spectroscopy, chemical composition, age, and environment of the galaxies. Second, for our purposes the important point may be this. It is remarkable that such a large population of galaxies can be usefully dissected. Keep in mind that each galaxy represents a complex system containing many stars and, to varying degrees, dark matter and interstellar dust.

Though our galaxy is just one among many, we need to discuss the local environment a little more specifically before we move on. In particular, to appreciate the magnitude of the extragalactic distance scale, we need to review the situation closer to home.

1.3.2 Our Immediate Neighborhood

To get a feel for astronomical distances, we will start small and then advance to the cosmological scale. The radius of our Earth is about 6 400 kilometers. As was known by the ancients using basic measurements and trigonometry, the Earth-Moon separation is about 60 Earth radii, or about 400 000 kilometers. The mean Earth-Sun

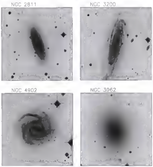


Figure 1-1 Four galaxies from the Automatic Plate Measuring survey. In these negative photographic images, stars within the galaxies are not individually resolved though the overall luminous structure of each galaxy is clearly indicated. Credit: Lahav et al. (1995)

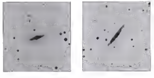


Figure 1-2 A side view of two galaxies from the Automatic Plate Measuring survey. Credit: Naim et al. (1995)

separation is about 150 million kilometers. Even these local distances can be hard to comprehend at a glance. The speed of light is a familiar, universal standard which will make our distance determinations a little more intuitive. Since the speed of light is 300 thousand kilometers per second, it takes light about 1.3 seconds to travel the 400,000 kilometers between the Earth and the Moon. So we refer to that distance as 1.3 light-seconds. Likewise, radiation from the Sun takes about 8.3 minutes to reach us here, so the Earth-Sun distance is 8.3 light-minutes. The major axis of Pluto's elliptical orbit is about 5.6 light-hours, which defines a rough size to the solar system. Since the speed of light is large compared to current rocket speeds, it is easy to see how it is that we are physically confined to our solar system for the time being.

One light-year, which is approximately equal to 10^{13} kilometers, proves to be a convenient astronomical unit. Though ten trillion kilometers seems quite large in Earthly terms, it is still easily imaginable using the familiar concept of parallax. There are hundreds of nearby stars whose distances can be measured using the parallax method (Mihalas & Binney 1981, Unsöld & Baschek 1991). Based on positional observations spanning six months we can use the diameter of the Earth s orbit about the Sun as a baseline for a simple trigonometric calculation Using this technique we can find the distance to our nearest neighbors, α Centauri and its dim companion Proxima Centauri, to be 4 3 light-years. There are about 50 such distance determinations for stars within about 15 light-years of us and this technique can be used with some success for a thousand or so stars within about 100 light-years Incidentally of these stars, some can be seen with the naked eye and some are too dim to see despite their proximity In astronomy we divide a circle into 360 degrees and 1 degree is further divided into 60 arc minutes and each arc minute is divided into 60 arcseconds Proxima Centauri shifts in angular position by 1 52 arcseconds, relative to the fixed celestial sphere of distant stars and galaxies over 6 months of observation If we take into account that the Earth has shifted by about 16 6 light-minutes during this period then we deduce the 4 3 light-year distance In addition to the light year, another frequently encountered astronomical length unit is the parsec which is about 3 26 light-years (an object one parsec distant would shift in apparent position by an arcsecond over three months) With the new orbiting telescopes and adaptive optics which avoid the smearing effect of the Earth s atmosphere, an angular resolution down to 0 001 arcseconds allows one to measure parallax out to a few thousand light years still well within the bounds of the Milky Way

There are other methods available to find distances Parallax methods can be extended by comparing measurements taken over a period of twenty years or more accounting for the solar system s orbital 220 kilometers per second motion about the galactic center (taking roughly 200 million years per orbit) As impressive as these important parallax methods are they still only probe our neighborhood within the Milky Way Other distance methods are generally variations on the following theme If we have two standard candles (astronomical objects which we believe to have equal intrinsic luminosity), and one is twice as distant as the other, and if there is no obscuring dust in the way, then the more distant standard candle will appear exactly one fourth as bright. This is an oversimplification, and such methods are generally quite complicated and imprecise in practice. One type of star in particular, known as a Cepheid, has been quite reliable as a standard candle. There are other standard candles of varying reliability, including certain types of stars, dust clouds, supernovae and galaxies. Over the years, such standard candle methods have become fairly robust within our galaxy and for neighboring galaxies. Concerning the extragalactic regions there are still discrepancies of a factor of two in these complex techniques.

Essentially all stars and galaxies which we observe are fixed on the celestial sphere, which means that we do not see them move at all, even though we repeatedly observe them over months and years. When one speaks of motions within our galaxy, and of the cosmic motions associated with large-scale structure, it is important to understand that such motions are impossible to see by trying to watch things move through a telescope. Effects giving rise to apparent motion, such as the Earth's rotation and the precession of the equinoxes, are accounted for when coordinates for a star or galaxy are specified. On Earth we specify a location of a city with two angles: the latitude (measured from the equator) and the longitude (measured from the Greenwich meridian). To specify the location of a star or galaxy, we imagine that there is an infinitely large celestial sphere about the Earth that has its own latitude and longitude lines of reference drawn on it. The celestial latitude is called the declination and is measured from the projection of the Earth's equator onto the celestial sphere. The celestial longitude is called the right ascension and is measured with reference to the Sun's celestial meridian at the vernal equinox (March 21) of some reference year, typically 1950 or 2000. So if you (1) know where your telescope is located, (2) can describe where your telescope is pointing, and (3) know what time it is, then you can unambiguously specify the astronomical coordinates of the star or galaxy in the center of the telescope's field of view.

1.3.3 The Milky Way and the Nearby Universe

The luminous disk of our galaxy is about 100 thousand light-years in diameter. Throughout much of the disk, its thickness is only about 2,000 light-years, although this increases at the central bulge. It is not a trivial task to describe our galaxy, mostly because it is quite complex, and the observational data are generally not simple to interpret. But we can still supply a broad view which will help orient us toward our later cosmological perspective (Mihalas & Binney 1981; Peebles 1993). As one might glean from Figures 1-1 and 1-2, the two main features of the Milky Way are the disk and the spheroidal component. The disk consists of two main populations. The main stellar population is smoothly distributed but becomes more concentrated toward the galactic center. These stars, like the Sun, are fairly aged (but are definitely not the oldest stars around) and lie close to the central galactic plane and travel in nearly circular orbits about the galactic center. The other disk population, associated with the spiral arms, contains a clumpy distribution of young stars, dust, gas, and nebulae. The chemical composition of the disk suggests that it is made of material that has already been processed by a previous generation of stars. The spheroidal component can be broken into a few sections. The galactic nucleus is the dense, violently active region within 10 light-years of the center. The galactic bulge is the central region about 20 thousand light years in diameter. The halo is the blob of matter, both luminous and dark, which envelops the entire disk, perhaps even extending beyond. Note that the halo is not actually evident in the photographs in Figures 1-1 and 1-2; only the increased central luminosity and bulge are clearly shown. The spheroidal component is mostly free of gas and dust, and its stars are generally quite old, roughly as old as the Milky Way itself. Some of these stars reside in globular clusters containing up to 100 000 stars. There are about 200 such globular clusters scattered about the inner galactic halo.

In Figure 1-3 we provide a rough schematic side view of the galaxy. Note that the familiar constellations and practically the entire night sky we observe with our naked eye are contained within a few hundred light-years of us. We try to indicate this on the figure, but keep in mind that in actuality this night sky region comprises a fraction of a percent the galactic volume. There are a few major exceptions to these visual limitations. To the casual observer strolling about the countryside on a clear night, the "milky way" is the faint, milky ribbon which encircles the sky. It is brightest toward the constellation Sagittarius, which is the direction to the galactic center. We are seeing the effect of distant stars within our own galactic disk, and so we get an inside panoramic view of our galaxy. Another important yet dim night sky object is the Andromeda nebula, which appears as a fuzzy star in the constellation Andromeda, near the square of Pegasus. Through a telescope (after a conventional photographic exposure of perhaps a quarter of an hour), this dim entity is revealed to be a giant spiral galaxy similar to the Milky Way. It is the closest significant galaxy to us, being only 2.5 million light-years distant.[3] The Large and Small Magellanic Clouds (seen only in the southern sky) are small irregular galaxies about 150 thousand light-years from us.

All galaxies, including our own, are dynamical systems which evolve throughout their lifespan. With such a large sample of galaxies available for observational study one might think that we would have ample opportunity to witness galaxies at various stages of evolution. Nearby galaxies are viewed as our contemporaries, as the light

---

[3] In the literature, e.g. Peebles (1993), this result is quoted with a 4% uncertainty. Even accurate distance determinations within our galaxy are uncertain by up to 10–20%. Other extragalactic distances are generally much more uncertain.

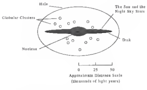

Halo
The Sun and the Night Sky Stars
Globular Clusters
Disk
Nucleus
0
25
50
Approximate Distance Scale
(thousands of light years)

Figure 1-3 Schematic side view of our galaxy

from them has taken only tens of millions of light-years to reach us. More distant galaxies are viewed well back in their history, as their light has been traveling hundreds of millions of light-years toward us. Although evolutionary effects have been noted and discussed in the literature, there are not many systematically useful results that are known. Lack of progress here is probably related to the lack of a specific theory of the galaxy formation process.

Nonetheless, we still know a fair amount about galaxies in general. For example, the disk material does not rotate together, like a rigid body. If that were the case, the stars and gas in the outer regions would have to travel faster in their trajectories to keep up with the inner material. What happens is that the outer matter (which has longer to travel) gets passed up by the inner matter. In fact, trajectories depend on the mass distribution within the galaxy, according to the equations of gravity. For example, in our solar system, the planets have speeds which decrease in inverse proportion to the square roots of their distances from the Sun, as explained by Newton

Such a velocity profile (known as a "rotation curve" when applied to galactic matter) is called Keplerian. In many galaxies (including our own) the rotation speeds are, roughly speaking, fairly uniform throughout the disk, so we would say these galaxies have flat rotation curves. For spiral galaxies, this is only part of the story, since the spiral arms are dynamically evolving density enhancements which rotate at a slower rate than the rest of the disk material.

Looking at galaxies from a cosmological perspective, an overall picture might be something like this. With our knowledge of the stellar populations, gaseous content and the dynamical information, we have a clue as to the age and evolutionary properties of galaxies. So we claim to know enough about galaxies to use them as tracers of the large-scale cosmic picture.

As our final point concerning galaxies, we mention the halo and dark matter, a matter which pertains directly to the large-scale structure problem. Returning to the solar system trajectories, we can say that the reason for the Keplerian decrease in planetary speeds is that the trajectories are governed by the large central solar mass. The other masses within the solar system only perturb this primary effect. We would not expect the orbital speeds within a galactic disk to decrease in a Keplerian fashion. However, as we neared the edge of the luminous disk, we would expect a characteristic decrease in the rotation speeds. With modern finely resolved measurements on dozens of galaxies, we are able to obtain spectrographic data spanning the various regions of individual galaxies. Such spectral data can be analyzed for Doppler shifts from which we obtain rotation speed profiles.[4] The rotation curves appear flat even beyond the luminous disk. If the trajectories are governed by conventional gravitational physics, then the emerging picture of a galaxy includes a large spheroidal blob (called the halo) containing dark matter and spanning some 100–600 thousand light-years.

---

[4] Shifts in the spectral pattern of galactic electromagnetic radiation are directly related to galactic motion—this is analogous to the acoustical Doppler shift whereby a musical pitch raises and lowers according the motion of the sound source. To find the velocity patterns within a galaxy, the measurements must be so precise as to show the differences in Doppler shift between various regions of the same galaxy.

We believe that all matter responds to the gravitational force in the same way. This statement would apply to regular, everyday matter (e.g. protons and neutrons) of course, but there is more to the story. The universal influence of gravitation would also include unusual matter (e.g. neutrinos, which have been known to the nuclear and particle physicists for years) and exotic matter (e.g. axions and other currently undetected particles). We know there is dark matter, we just do not know what it is made of. The exotic forms of dark matter are often called "nonbaryonic," as "baryonic" is a term loosely applied to protons and neutrons (which are the heaviest form of regular, stable matter). Galactic halos may be made from one type or from a mix of different types of dark matter. The issue is further complicated by the fact that dark matter is detected on larger, cosmological scales. There are arguments that some of the dark matter must be nonbaryonic. These dark matter issues seem to span a vast range of distances. But the galactic scales are almost certainly related to the larger cosmological scales. Recall that our galaxy, as well as all the other galaxies and even larger structures, did not exist far into the past. They were formed within an expanding universe that was much smoother (less lumpy) in the past. The dark matter connection is simply the fact that all these structures were formed as a result of gravitational instability. In smaller regions ranging from stellar to galactic halo scales, electrodynamics and gas dynamics play crucial roles which further complicate the situation.

1.4 Structures in an Expanding Universe

1.4.1 The Big Bang Model

In the previous sections, we were trying to find a perspective from which we could appreciate the nature of extragalactic data, and so now we turn to the larger picture of the big bang and cosmic structure.

All distant galaxies are receding from us. Each galaxy has a well defined, measurable Doppler shift in its electromagnetic spectrum. If there were no relative motion between us and some particular galaxy, then we would detect no Doppler shift for it. However, essentially all galaxies have spectra which are shifted toward longer wavelengths. In the visible region of the spectrum, red has the longest wavelength and blue has the shortest wavelength. Technically speaking, we say that all galaxies have positive redshifts, which means that all galaxies are traveling away from us. This is analogous to the following Earthly experiment. If we distributed identical tuning forks to various people and then asked them to move about, we could analyze their motions from the various pitches we heard. If all notes were received at a lower pitch than the standard, that would mean that all the people were traveling away. Moreover, the faster the recession speed, the lower the pitch we would detect. Similarly for galaxies, we say that the amount of redshift is proportional to the recession velocity of the galaxy.

If we trace the expansion backward in time, then galaxies used to be closer together as we imagine the universe 5 billion and even 10 billion years ago. We can go a bit further back, but then we are not really talking about galaxies—rather we have oddly shaped galactic blobs and, slightly before that, protogalaxies and first generation stars. We can even go much further back than this, still within the realm of a consistent scientific framework. In doing so we are simply applying standard physical laws in an extrapolated environment. This background environment is not just the idle choice of theorists: the background obeys the laws of general relativity, thermodynamics and particle physics. For example, the ambient temperature of our observable universe increases as we probe further into the past, because in the past the mean density was larger than it is now (the same matter used to occupy a smaller volume). We can get a similar effect in a laboratory. If we compress some gas in a thermos (to avoid heat exchange with the environment), the gas temperature increases. Also, if a gas expands without heat exchange, then the temperature lowers (e.g. the material expanding out of a fire extinguisher is colder than its storage temperature). In the big bang model, as the universe expands, its ambient temperature decreases. By analyzing the faint but ever present cosmic background radiation, we find that the current temperature of the universe is about 2.7 K, i.e. 2.7 degrees above absolute zero.[5]

The nature of the expanding universe was revealed to the world in the 1920s and 1930s following the pioneering observations of Edwin Hubble. First, he was able to resolve individual Cepheid stars in the Andromeda nebula (as it was then called) and thus determine its distance to be far outside the borders of the Milky Way. The true universe of galaxies was discovered. Furthermore, by the time Hubble was through he (and others, primarily Milton Humason) had measured the Doppler shifts of over one hundred galaxies. All but a few neighboring galaxies were receding from us. In a statistical sense, his data were pretty sparse, but he was able to conclude that in general the recession velocities are in direct proportion to the distances. This is now called Hubble's law. The observations were statistically similar in all directions which agrees with the postulate that we live in a homogeneous and isotropic universe. It was this important concept of homogeneity and isotropy that led Einstein (1917), Friedmann (1922), and Lemaitre (1927) to the fact that in general relativity the only physically possible universes are those which expand.[6] Lemaitre pointed out the basic prediction that the galaxy distances and recession velocities are proportional, and the redshift data at the time were starting to hint at this. Hubble's (1929) paper "A Relation Between Distance and Radial Velocity Among Extra-galactic Nebulae," was conservative, scientifically sound, and left little room for doubt.

---

[5] The Kelvin (K) temperature scale is the same as the Celsius scale shifted so as to begin at absolute zero. For example, water freezes at 273 K, room temperature is about 300 K, and water boils at 373 K. The surface temperature of the Sun is about 6000 K, and hot x-ray emitting gas in some galaxy clusters may reach temperatures of 10^8 K (100 million degrees).

The expansion of the universe corresponds to the continual stretching of the actual fabric of space. This can be difficult to imagine, but it is a basic consequence of general relativity—a theory which purports to specify the fabric of spacetime based on the matter contained therein. We might be tempted to think of the big bang as an explosion which sent all the matter shooting outward into an empty space. But that is not the case. A standard visual aid is to imagine the universe as a large loaf of raisin bread, where each raisin represents a galaxy. The raisins get farther apart as the baking bread expands. The individual raisins stay the same size—galaxies are gravitationally bound systems which do not expand along with the overall spatial fabric. Pretend that the Milky Way is one of the raisins (it does not matter which one we pick). From the point of view of that home Raisin, the others are all traveling outward. Let us say that the loaf doubles in size in one hour. Then a speck that was initially 1 cm from us doubles its distance to 2 cm in one hour. But during the same hour, another speck which started out 2 cm from us ends up 4 cm distant, as all distances within the yeasty fabric have doubled. This means that the speck twice as distant traveled away at twice the speed, in analogy with Hubble's law. The blatant flaw in this bread baking analogy is that the bread was expanding into something. Because general relativity is describing the most primitive of entities, spacetime itself, the entire picture can be difficult to imagine. The big bang happened everywhere; no one place is the origin of it all. Present astronomical observations are consistent with an infinite universe, which may further confound one's imagination.

---

[6] Einstein's brilliant paper of 1917 introduced and rejected the concept of an evolving universe—at the time he wished for a static universe. But his paper was written from a purely theoretical standpoint. As modern data such as the redshifts later came pouring in, the assumption of an unchanging universe had to go. The Russian meteorologist/mathematician Friedmann took the reality of the expansion seriously. The Belgian cleric/astrophysicist Lemaitre began developing early versions of a big bang theory.

These ideas, as difficult as they are to visualize, can be tested. For example, the particular way in which the galaxies in a sample become dimmer with distance can be used as a test as to determine whether the apparent recessions are due to an expanding spacetime metric or some other unusual effect. Only recently have the data become sufficiently precise so as to verify the reality of the expanding metric in this way.[7] But such verification came with little fanfare, for the cosmic background radiation discovered in the 1960s had provided data which can only be explained by the expanding, cooling spacetime model. Perhaps there is an easier way to come to terms with the expansion phenomenon, using our knowledge of the speed of light. Since the recession velocity is proportional to distance, some of the most distant galaxies we observe have recession velocities in excess of the speed of light. But does not that violate the cosmic speed limit, a foundation of Einstein's theories? Not at all, for even two widely separated galaxies, apparently moving relative to each other at super light speeds, are simply sitting there, approximately at rest in their neighborhoods of space. The apparent recession speed is only the result of the space between expanding—by the laws of general relativity, the wavelength of traveling light expands along with the space, and this change of wavelength is perceived as a Doppler shift If a galaxy were moving away with an actual physical velocity faster than the speed of light we would never be able to observe the light from it On the other hand if all galaxies were just sitting there there would be no structure—matter has to move around in order to become clustered and such motion is generally a small but important addition to the overall Hubble flow of galaxies

---

[7] Hubble's protege Alan Sandage recently published an account of this surface brightness test and other fundamental considerations in an article entitled "Evidence That the Expansion Is Real" (Sandage 1992).

The cosmic background radiation is the earliest direct relic we can observe from a hot universe The radiation we are now receiving was emitted when the temperature (as defined by the thermal bath of electromagnetic radiation) was about 3000 K, some 10 billion years ago just a few hundred thousand years after the expansion of space began To get a handle on this topic we have to trace backward to within minutes after the beginning of expansion As we ponder the state of the universe at various stages back to several billion years into the past, we at least have recourse to a certain amount of basic intuition For during this time galaxies existed in various states of evolution, and some of the matter was flowing smoothly into structures The ambient temperature at the stage of galaxy formation was a cool 100 K or so But as we probe deeper into the past macroscopic structures cease to exist, and we enter the realm of atomic physics with only the two lightest elements hydrogen and helium and some radiation in appreciable abundance, and temperatures are in the thousands of degrees the universe is a hot gas Earlier on the order of a second after the big bang temperatures are on the order of billions of degrees, and the fundamental processes of particle physics rule the cosmic soup of protons neutrons electrons and neutrinos (and their antiparticles) and photons (the quantum of electromagnetic radiation)

When all physical characteristics of a radiating system except the ambient temperature have been washed out generally because of multiple interactions or "thermalization," we say the system emits blackbody radiation of a certain temperature

In the early 1900s Planck Einstein and others demonstrated that light (i e electromagnetic radiation) is quantized into particles which carry energy in inverse proportion to the wavelength of the light[8] A blackbody gives off radiation at all different wavelengths, but the wavelengths are distributed with a spectrum (called the Planck distribution) uniquely defined by the temperature For example the Sun s thermalized outer regions emit an approximate blackbody spectrum corresponding to about 6000 K with all wavelengths represented but with the peak intensity lying in the visible region (wavelengths of about a hundredth of a millimeter) The cosmic background radiation is distributed exactly according to a Planck distribution corresponding to a temperature of 2 7 K with peak intensity lying in the microwave region (at millimeter wavelengths)

So from all directions of the sky we are being bombarded with a very uniform low energy background radiation corresponding to about 400 million photons per cubic meter, detectable only by sensitive instruments. What is the density of regular matter in our part of the universe? Roughly speaking just by counting the thousands of galaxies near us (the mean separation between galaxies being some 15 million light-years) we can estimate the average baryon (neutron or proton) density to be on the order of one per cubic meter Although the photons far outnumber the baryons, presently the radiation contributes a negligible portion of the overall energy in the universe (the jargon presently the universe is "matter dominated) Far into the past, say for the first 10 thousand years or so when temperatures were on the order of thousands or millions of degrees, the energy of radiation comprised most of the universe's energy (the universe was "radiation dominated") These photons have traveled long and far to reach us and they began their journey during an earlier hotter epoch

---

[8] So an x ray photon carries more energy than a visible light photon which carries more energy than a microwave photon etc

By number, baryons are presently about a billion times less populous than the cosmic photons, but the baryons are now playing the crucial role in our universe. From complex astronomical analyses, we detect that ordinary matter is presently made up of about three quarters hydrogen and one quarter helium, and such abundances are one of the great predictions of the big bang model. When the temperatures are on the order of 10^{10} K, less than a second after the universal expansion began, protons and neutrons could not even consider joining up to form stable atomic nuclei. A little bit later, within the first three minutes or so, simple nuclear reactions were able to transpire. But the nature of the interactions—involving only the simplest combinations of small numbers of protons and neutrons—and the cooling and dilution of the universe allowed only a brief opportunity for nucleosynthesis. Difficult calculations, first performed in the 1960s, provided unambiguous predictions concerning elemental abundances; the success of these predictions is, in a sense, an even deeper probe back into time than the cosmic background radiation. As an important note concerning galaxies and humanity, we point out that although the present baryon population is primarily hydrogen and helium, about 2% of regular cosmic matter is comprised of heavier elements such as carbon and oxygen. But these important elements were not constructed during the early universe; they were synthesized later in stellar nuclear reactions.

1.4.2 From Early Microscopic Processes to the Recent Formation of Galaxies

Many of the events of the hot, early universe have had great influence on the present. We will just touch on a few things, primarily to emphasize the importance of the hydrogen and helium synthesis and to portray the atmosphere leading to the epoch of structure formation.

Throughout much of the history of the early universe, matter and radiation were in approximate thermal equilibrium. For example, less than a second after the big bang, the universe was merely a hot (over 10^{10} K) gas of particles which was quite dense (over a million times denser than ordinary water) and behaved like a sea of radiation. These particles were thermalized because of the many interactions between the photons, electrons/antielectrons, neutrinos/antineutrinos, and the very small but important population of neutrons and protons.

To understand these early times, we need to keep in mind the equivalence of mass and energy. Mass is the source of gravitation, but since Einstein's time, our interpretation of this has become more general. Any lump of matter or pure energy (e.g. electromagnetic radiation or even more exotic types of energy) participates in gravitational forces. The universe is expanding, but the expansion is getting ever slower. In observations, we cannot see this directly (we would have to wait billions of years), but the data imply it. The reason for this deceleration is similar to the reason that a rock thrown upward from Earth slows down (note that the rock eventually comes back down unless it has reached the Earth's escape velocity). The 400 million photons per cubic meter (at 2.7 K) contribute a much smaller energy density compared to the approximately 1 baryon per cubic meter. So the baryons have a much larger effect on slowing the expansion of space, according to the general theory of relativity. In the early universe, various forms of hot radiation possessed sufficient energy density to be a primary factor in the cosmic deceleration. During the radiation dominated era, photons were by far the most populous species, accompanied by a smaller amount of neutrino radiation and a small but crucial population of electrons and light nuclei (hydrogen and helium). This era lasted about 10 thousand years, starting just a few minutes after the big bang. In older literature, the early universe is sometimes referred to as the primeval fireball, and this is a fitting description of a very different era.

At an age of about a tenth of a second, the universe consisted of particles in equilibrium, zipping around at the speed of light. Conceptually, this environment is fairly simple. Only a small part of the huge zoo of elementary particles is present at this stage of the universe. A temperature of more than 10^{10} K means that the particles are interacting at high energies—energies which are larger than the rest energy of an electron/antielectron pair. This means that electron/antielectron pairs can be created and destroyed at will during fundamental interactions involving photons. This is an example of Einstein's mass-energy equivalence at work. But the energies are relatively low in the sense that other elementary particles are for the most part not in existence, as they have higher threshold energies (i.e. they are more massive). We know that a population of neutrinos exists at this time, as they are a stable, practically massless, particle. Although neutrinos are weakly interacting, at these densities they are thermalized with the other particles.

Last but not least, we have the protons and neutrons. The neutrons are not stable (they decay in minutes on Earth), but in this environment, elementary interactions can continually interchange the protons and neutrons.[9] A priori, we might not have reason to include the baryons in the picture, but the empirical evidence (the current photon to baryon ratio of about a billion to one, the predominance of matter over antimatter, and the fact that we exist) says that we should. Neutrons are slightly more massive than protons, so as the universe expands during the initial few minutes, it becomes energetically favorable for the baryon reactions to lead to an excess of protons over neutrons. By this time, the temperature is low enough that electron/antielectron pairs cannot be created, so their population is decimated as they decay into photons.

---

[9] For example, we have the processes $p + \nu \longleftrightarrow n + e^+$ and $n + \nu \longleftrightarrow p + e^-$

Protons and neutrons which were previously unable to bond are able to combine but the helium nucleus (two protons and two neutrons) is the only stable combination synthesized to an appreciable amount So after a few minutes of universal expansion the baryons are in the form of protons (i e hydrogen nuclei) and helium nuclei The excess of protons over neutrons leads to a robust prediction of about 75% hydrogen and 25% helium (there is a leeway of several percent in the calculations)

The above examples illustrate a certain type of early universe phenomenon Stages of approximate equilibrium are interrupted by important deviations from equilibrium The reduction of the neutron population and the freezing out of the electron/antielectron pairs cited above are fairly typical examples Generally there is a tension between the rates of fundamental processes and the rate of the Hubble expansion during the early universe The rates of particle interaction depend upon the density of particles the strength of interactions, and the particle energies whereas the spatial expansion dilutes the particle population and lowers the energies of interaction among particles In a qualitative sense we can see why there was only a brief opportunity for atomic nuclei to form early on things were too hot and dense, later on the environment was too dilute

As odd as it seems the events of the early universe were changing by the millisecond or by the minute Today one would have to wait billions of years to see change of such magnitude During the early times gravity was important only in the most global sense—the energy density of particles dictated the expansion rate of the universe It was the microscopic interactions of elementary particles which proved most influential

Today as we look at the matter beyond our galaxy, we see a fair amount of structure. But if we gloss over the structure, we see a universe which seems somewhat homogeneous overall In the far past (e g. during the era of nucleosynthesis) the universe was very uniform, however. The overall average density of the universe has been continually decreasing. After nucleosynthesis, the mean density (of the radiation) was on the order of 10 grams per cubic centimeter, or ten times that of water. Currently the overall cosmic density is about 10^{-30} grams per cubic centimeter (or, as we have been saying, about one hydrogen atom per cubic meter). As far as we can tell, there have always been fluctuations about the average density, and these fluctuations have grown over time because of the force of gravity. Presently, we can perform galaxy counts in a hypothetical cube of length 20 million light-years per side and compare the number of galaxies contained in the cube as we move it around the sky. We find that there are significant fluctuations in the galaxy counts in such an experiment. If we could possibly perform this same experiment some 10 billion years ago, we would find that the galaxy counts would have smaller variation—i.e. the density fluctuations have grown over time. In considering such density experiments, it is only meaningful to consider cosmic scales, as the local densities in our solar system or galaxy are huge compared to the mean large-scale density. The intricacies of our solar system or the spiral arms of a galaxy are considered to be fine details; they are independent systems which are not directly related to the overall evolution of cosmic fluctuations. This is just another way of saying that the large-scale structure problem deals with regions larger than galaxies.

The executive summary would sound like this: The universe has an average density which steadily decreases due to the expansion of space. Density fluctuations have been amplified due to gravitational instability, and the growth of these fluctuations has led to the structure which we observe today.

The universe was highly uniform early on, but already at that time there were small density fluctuations. Naively, we might think that areas of slightly higher density would attract matter to form lumps which could further attract matter, because of the force of gravitational attraction. And the slightly underdense regions would lose matter which might flow toward the clustered matter. This idea is not too far off the mark. Isaac Newton considered the point in a cursory fashion three hundred years ago. Early in this century, James Jeans made some quantitative estimates concerning the way in which a large-scale (astronomical) fluid might break into lumps via gravitational instability. This work could be updated to cosmic proportions, and in 1946 E.M. Lifshitz made some headway within the context of the general theory of relativity. The modern era in cosmological structure formation analysis was primarily led by Ya.B. Zel'dovich in Russia and P.J.E. Peebles in the United States, beginning in the 1960s. Largely through their leadership, a picture has emerged over the last thirty years which may yet gel into a coherent theory. One remarkable point, which has added much excitement within the scientific community, is that the theory and the observational data have both been improving steadily over this period.

Consider the universe when it was much smaller and hotter, say sometime after nucleosynthesis, during the radiation dominated era. Density fluctuations were already present, but they were not yet able to grow, because of the stultifying effects of radiation pressure. The gravitational properties of the universe would be determined by the gas of radiation. The radiation, like any form of mass or energy, would feel the urges of gravitation, but it also would feel the effects of its own pressure forces, much like a familiar fluid. The effects of fluctuations would be reduced to a minimal sloshing around of the photons and baryons, a photon gas version of a sound wave. Any nonbaryonic dark matter present would follow the lead of the radiation. However, after about 10 thousand years of radiation domination, the temperature will have cooled to the extent that matter comprises a larger component of total energy density. Matter fluctuations are free to grow. Thus began the present era, known as the matter dominated era.

The last important event before structure formation came into full swing occurred about 300 thousand years after the big bang. At this time the universe consisted of photons (which no longer dominated), neutrinos (which had been freely streaming since the era of nucleosynthesis due to their very weak interactions), hydrogen and helium nuclei (positively charged, as the universe had always been too hot for electrons to join with them into neutral atoms), and electrons (negatively charged light weight particles waiting to combine with the nuclei). At about the 300 thousand year mark, the temperature had lowered to about 3000 K, low enough for the electrons to safely combine with the nuclei, thus forming neutral atoms. This transitional period is traditionally called "decoupling." Before this, matter and radiation were closely coupled (and in fact were at the same temperature) because of the frequent photon-electron interactions. After the neutral atoms form, these scattering interactions are effectively silenced. At this point, the universe is transparent to the hot photons, and they travel freely for the rest of time. This is the faint radiation we are presently detecting from all directions of the sky. The photons were thermalized early in the universe, and so the radiation emitted during the decoupling period was blackbody radiation corresponding to a temperature of about 3000 K. The expansion of space in the intervening 10 billion years or so has simply redshifted the light to longer wavelengths, corresponding to a photon temperature of 2.7 K.

The final piece in the puzzle which authenticates the paradigm of big bang structure formation was discovered by NASA's COBE satellite in 1992.[10] Very soon after entering orbit and starting operations, the satellite observatory was able to dramatically update the accuracy to which we knew that the cosmic background radiation was exactly a blackbody radiation. After some more years of data taking and processing, COBE was able to detect slight fluctuations in temperature (1 part in 100 000) between various parts of the sky. Recall that the fluctuations in matter density had started to grow ever so slightly by the time decoupling occurred. These small fluctuations left a statistical imprint on the sea of radiation just before it decoupled. Cosmologists had been expecting the experimental physics groups to detect these temperature fluctuations for many years. But the measurements had not been quite sensitive enough. The data are quite impressive in the sense that the experimenters detected some cosmic noise superimposed onto their detector's instrumental noise. There was a big hubbub in the news media at this time (April, 1992) because this event demonstrated the existence of some sort of statistical structure in the fluctuations of the truly ancient cosmic radiation.

---

[10] COBE stands for COsmic Background Explorer.

The question naturally arises: Where did these fluctuations originally come from? One very conservative approach is to simply sidestep the question. In order to provide more focus, we narrow the discussion so as to consider just the growth of fluctuations since decoupling, and we treat the statistical characterization of the early fluctuations as initial conditions. This is a standard approach. Another approach, common since the early 1980s, has been to extrapolate the laws of physics back into the very, very early universe in order to seek a deeper understanding of the basic initial conditions of the big bang model. The large-scale structure problem in the simplest sense is concerned only with the growth of small fluctuation into present day structures. Such growth transpires via gravitational instability only after the universe has entered the matter dominated era. The COBE satellite data provide us with a direct link to these fluctuations. That is why the decoupling transition provides a natural boundary for the large-scale structure problem.

1.4.3 Structures

It is not possible to get a picture of large-scale structure by simply looking through a telescope. The distribution of matter has to be mapped out piece by piece. Such mapping has been going on for over fifty years, and only during the last twenty years or so has a reasonable picture of cosmic structure begun to accumulate in the data.

There are about 20 galaxies within about 15 million light-years of us. This cloud of loosely knit galaxies can be thought of as a somewhat isolated gravitational system, called the local group. The Milky Way and Andromeda are its two largest galaxies; the other galaxies are merely satellites or dwarfs. About half of these galaxies (including Andromeda) possess blueshifts rather than redshifts, because the Doppler shift is mostly showing their local movements rather than the Hubble flow. If we do not consider our neighboring galaxies, then we can say that all galaxies for which we have Doppler measurements are redshifted, which means that they all are at least approximately following the outward Hubble flow.

Beyond the local group, many clusters containing 10–100 galaxies have been observed. During the 1950s, the Palomar Sky Survey data were used to catalogue thousands of relatively bright clusters.[11] These clusters were indicative of a lumpy overall distribution of galaxies, but the data were still too sparse to draw definite conclusions as to the nature of clustering. Interestingly, even in the 1930s, as such structures were first being observed, some astronomers (most notably Fritz Zwicky) were suggesting that a bit of extra dark matter must be present to gravitationally bind these clusters.

Also during the 1950s, it was proposed that we, along with perhaps thousands of nearby galaxies, were part of a so-called local supercluster. This innovative idea was formulated by de Vaucouleurs (1953), but the idea of such a large structure met with some resistance, and only in the late 1970s were skeptics finally convinced that the local supercluster really existed. It was difficult to extract structural information from the data before the 1970s, largely because redshift measurements were difficult and therefore scarce. As it turned out, a large, reliable data base of galactic redshifts would be a crucial element in the mapping of cosmic structure.

---

[11] The standard references are to Abell (1958) and Zwicky et al. (1961–68).

To describe the 3-dimensional distribution of galaxies in space, we need detailed information on the position of each galaxy. If we specify the right ascension and declination, we have only pinpointed a galaxy's location on the celestial sphere. To complete the picture, we need to know its distance from us. Direct distance determinations using standard candles are few and far between, and as we noted earlier, the numbers obtained are still a bit unreliable. Recall that Hubble's law says that the recession velocity of a galaxy is proportional to its distance from us. This is the saving grace of observational cosmology, for we are able to use the measured recession velocity as an approximate indicator of distance. Over the years, observational astronomers—who are an interesting mix of theoreticians and experimental physicists—have been able to provide newer and faster techniques for various measurements, including fundamental redshift determinations. Fancy electronics and computers have been the norm for some time now in astronomical data acquisition, and since the mid 1980s improvements in the redshift data base have been significant.

If we have angular positions and redshifts for a sample of galaxies, then we are able to formulate a picture of their spatial distribution. The redshift is a direct measurement of the recession velocity in the following sense. If, for example, the pattern of spectral lines for a particular galaxy is shifted by 1%, then its apparent recession velocity is 3,000 kilometers per second (which is 1% of the speed of light)

The formula which expresses Hubble's law is

recession velocity = Hubble constant × distance.

and the current value of the Hubble constant seems to be about 20 kilometers per second per million light-years. This value is currently uncertain by 50% or more, but many astronomers believe that a value will be more firmly pinned down by future Hubble Space Telescope measurements. But we still know the relative picture: if one galaxy has three times the redshift of a second galaxy, then the first is approximately three times as distant as the second. This is sufficient information to map the 3-dimensional structure of the galaxy sample. To take our previous example, a galaxy with a recession velocity of 3,000 kilometers per second (1% redshift) lies at a distance of about 150 million light years using the current value of the Hubble constant. When jargon is being used freely, astronomers will often interchange the phrases "recession velocity," "redshift," and "distance."

By the late 1970s, a few thousand galactic redshifts had been measured, and astronomers were able to confirm that our local group is one component of a rather massive structure. There are about 3000 galaxies within about 140 million light-years of us, and most of these comprise a somewhat flattened blob of galaxies called the local supercluster. Looking at modern presentations, such as Tully and Fisher's "Nearby Galaxy Atlas,"[12] which display maps showing each galaxy on a coordinate grid, we notice that the galaxies form a definite structure, something like a diffuse pancake. A simple illustration, from Tully (1982), displays some of the key features of the nearby galaxy distribution; this is reproduced in Figure 1-4. There are 2,175 dots in Figure 1-4. Each dot represents an entire galaxy which was observed to be

---

[12] This fascinating coffee table book (Tully & Fisher 1987) displays the location of about 3000 nearby galaxies, utilizing both their positions on the celestial sphere and their redshifts. The clarity of presentation and the colorful graphics presented in this atlas uphold the tradition of the finest Earthly atlases.

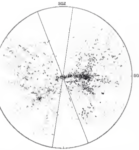


Figure 1-4 A view of the local supercluster. Each dot represents a galaxy. Our galaxy, if pictured, would be a dot at the center of the circle. The dots very near the center of the diagram are in our local group of galaxies. The flattened pancake-like structure dominating the center of the picture looks string-like, because we are viewing it edge on. Most of the 2 175 galaxies shown here are physically related to the local supercluster. The diameter of the circle within which all these galaxies lie is about 280 million light-years. Credit: Tully (1982)

"nearby," i.e. within 140 million light years of us and relatively bright. The figure is drawn from the perspective of a far away observer who sees the local supercluster pancake from an edge on view; the pancake edge is horizontal in this figure. As a technical note, we point out that in Figure 1-4 the dots are plotted using the SGY and SGZ coordinates from de Vaucouleurs' super galactic X,Y,Z coordinate system. These coordinates were specifically chosen so that the SGZ axis would be perpendicular to the pancake. On the lower right part of the pancake, the Virgo Cluster, the nearest significant cluster of galaxies, is apparent.

Figure 1-4 illustrates an important aspect of our view of the universe. Note that there are few dots plotted in the two 30 degree wedges in the upper and lower parts of the figure. This is because we make astronomical observations from within the thin disk of our Milky Way, which contains some obscuring dust. Within about 15 degrees of the galactic plane, there is essentially a zone of avoidance for extragalactic observations.

What do things look like when we probe far beyond the local supercluster? We can get an idea of what the very distant universe looks like by examining a few images obtained and processed by the electronic instruments associated with the Hubble Space Telescope. Perusing such images may help us appreciate the nature of deep space data. But to get a feel for the nature of large-scale structure we will look at maps which were produced from the compilation and analysis of very many astronomical images.

When we analyze stars or galaxies through a telescope, we need a quantitative way to characterize how bright or dim the objects appear to us. To do this, we customarily extend the system developed by the ancient Greeks, who divided the night sky stars into six classes of brightness. The brightest stars were classified as first magnitude and classes were equally divided down to the dimmest stars that could be seen with the naked eye and those were dubbed 6th magnitude In modern times we keep track of the amount of electromagnetic radiation deposited on the telescope's photographic plates or electronic detectors Because of the great light gathering capabilities of telescopes and because modern instruments are sensitive to levels of radiation millions of times dimmer than humanly detectable we may see deep faint images of distant galaxies Human eyes are logarithmically sensitive, and it turns out that the six magnitudes of brightness (which we still call 1st through 6th) which describe nearly all the night sky stars, actually correspond to electronically detected radiation energies spanning a factor of 100 We continue this system in the sense that the higher the magnitude of a star or galaxy the dimmer is the radiation we receive and an increase of 5 magnitudes always corresponds to a decrease by a factor of 100 in the amount of radiation received The thousands of galaxies in the local supercluster are too dim to see with the naked eye Most of them have apparent brightnesses somewhere in the range of 9th to 14th magnitude (note that an 11th magnitude galaxy is a hundred times too dim to see with the unaided eye) The Automatic Plate Measuring (APM) galaxy survey contains about 2 million galaxies brighter than the 19th magnitude, and therefore this survey probes far beyond the local supercluster In fact, the APM survey can detect a galaxy more than 1000 million light-years distant

The Coma cluster of galaxies is considered to be a fairly typical large, regular cluster This cluster is so named because all of its galaxies appear near the constellation Coma Berinces on the celestial sphere There are a few thousand galaxies with similar redshifts and angular coordinates which would seem to indicate that they all form a fairly tight group Further analyses of the system (e g dynamical analyses independent distance determinations) confirms this The Coma cluster galaxies have recession velocities near 7000 kilometers per second which means that they are all

Figure 1-5 Gazing beyond the Coma Cluster using the Hubble Space Telescope. The large image near the center (elliptical galaxy NGC 4881) and the spiral galaxy to its right reside on the outskirts of the Coma Cluster, some 330 million light-years away. The Coma Cluster contains hundreds of galaxies, but this narrow picture shows just two of its members. A few of the dots shown are foreground stars in our Milky Way. Most of the remaining objects are galaxies far beyond the Coma Cluster. Credit: Hubble Space Telescope WFPC Team and NASA (1995), Space Telescope Science Institute press release ST ScI OPO PR95-07

about 350 million light-years distant The galaxies within the core of the cluster lie within one degree of each other on the celestial sphere

As a bit of galactic trivia, we note that the core of the Coma cluster lies very near the North Galactic Pole Our navigational North Star Polaris lies near the North Celestial Pole defined by the Earth's spin axis The North Galactic Pole is analogously defined by the Milky Way's spin axis Because the Earth's spin axis wobbles (the axis completes one precession every 25,000 years) Polaris will not be the best "North Star" in a few thousand years. But even a million years from now, the Coma cluster will still be an excellent indicator of the North Galactic Pole

The recently repaired Hubble Space Telescope (HST) has provided some of the finest deep images yet obtained We display two HST images drawn from recent press releases The first Figure 1 5 shows a clear image of two Coma cluster galaxies (350 million light-years distant) along with some background galaxies far beyond Figure 1 6 shows a sequence of state of the art faint images of distant galaxies, some of them billions of light-years distant

In Figure 1-5, we see a giant elliptical 13th magnitude galaxy, known as NGC 4881, in the center of the image To its right is a spiral, 16th magnitude galaxy Both of these are members of the Coma Cluster although they reside about a third of a degree from the cluster center The image is fairly narrow only a few arc minutes across (NGC 4881 has an angular diameter of about 1 arc minute) and so does not include the many other cluster members Aside from a few foreground stars (from our Milky Way), the other objects in the background are galaxies far beyond the Coma Cluster The figure has an odd shape since it is actually a graft of four images, three large squares and one smaller square in the center from the three Wide Field Cameras and the one Planetary Camera aboard the orbiting telescope Even with the very sensitive electronics used this image required 16 exposures, each lasting 15 minutes

In a couple of the large-scale maps we will show shortly (Figures 1-7 and 1-8), the Coma Cluster will appear in its entirety.

To compare this image with the six galaxies previously shown in Figures 1-1 and 1-2, note that those six galaxies were of the 12th or 13th magnitude, were three to five times closer than the Coma Cluster, and had angular diameters of around 3 arc minutes.

As we look at ever more distant objects, we are probing further into the past. Nearby galaxies (closer than, say, 100 million light-years) are viewed as our contemporaries, as the light we receive from them was emitted only 100 million years ago. Galaxies which are 1000 million light-years distant (meaning that we are receiving light emitted a billion years ago) appear to be similar to nearby galaxies, although there can be signs of galactic evolution in their spectroscopic data. The first real visual impression of galactic evolution was seen in the high resolution images of faint distant galaxies obtained by the Hubble Space Telescope, as shown in Figure 1-6 from an HST press release (Dressler et al. 1994).

As we go from left to right in Figure 1-6, each column shows images from successively older epochs. The two galaxies in the left hand column of Figure 1-6 are nearby. Their recession velocities are less than one percent of the speed of light, and so they are nearby and represent our contemporaries. The three galaxies in the second column have recession velocities near 30% of the speed of light, and we are viewing them when the universe was about two thirds its present age. The third column shows galaxies when the universe was about one third its present age; the recession speeds are roughly equal to the speed of light. The last column contains galaxies with recession velocities over 300% of the speed of light; these are some of the most distant galaxies currently resolvable. The distinction between ellipticals and spirals becomes quite blurry, although the image at the top right has the spectroscopic profile of a

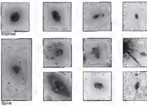


Figure 1-6 Galaxy evolution as revealed by the Hubble Space Telescope The four galaxies in the top row are classified as ellipticals, the seven below as spirals The two galaxies shown in the far left column are relatively nearby (within 30 million light-years) and are characteristic of the present epoch. The galaxies in each successive column to the right are far more distant and so existed in a younger universe The far right column shows newly formed galaxies when the universe was one tenth its present age The previous Figure 1-5 and particularly the last three columns in the figure above represent the state of the art in distant galaxy observations attainable after the December, 1993 repair of the orbiting telescope Credit Dressler Dickinson Macchetto, Giavalisco and NASA (1994), Space Telescope Science Institute press release STScI PR94-52C

mature elliptical Analyses of deep images are beginning to reveal some of the secrets of how galaxies form evolve and even merge with each other

1.5 Mapping the Large-scale Structure

In order to make headway into the large-scale structure problem we do not have to probe nearly as far back as the faint images in Figure 1-6 The current maps of the large-scale galaxy distribution—we show a few such maps below—provide a significant though still incomplete, glimpse of the large-scale structure picture The COBE cosmic background radiation fluctuation data provide important constraints on the clumpiness of the ancient universe Since radiation and matter were coupled in the radiation era of the early universe the smallness of the background fluctuations detected by COBE implies that the distribution of matter was not very clumpy If we consider both the current galaxy maps and the constraints provided by the cosmic background radiation data then it would appear that a reasonably fair or statistically representative sample of galaxies can be obtained using galaxies within about 1500 million light-years (i e with recession velocities less than about 10% of the speed of light) While there are millions of galaxies this close to us, it is a technically difficult matter to obtain large samples of galaxies where the samples are in any sense complete or thorough

The first large-scale mapping project was carried out by Shane and Wirtanen during the 1960s By visually analyzing photographic plates from the Lick Observatory, they were able to tabulate the angular positions of about a million galaxies on the celestial sphere The survey plates which they inspected included galaxies down to a dimness of about the 18th magnitude. Galaxies come in various sizes and intrinsic luminosities—we say "intrinsic" to distinguish the luminosity (how much radiation a galaxy actually emits) from the directly observed apparent brightnesses (how much radiation we detect far away) Since we have some statistical information concerning the distribution of galactic luminosities we can fill in the scientific details of the following truism surveys which include dimmer galaxies are probing deeper into the universe. Since the Lick catalog includes galaxies down the the 18th magnitude we can estimate that it contains galaxies out to a depth of about 900 million light-years

To collect the data for the Lick Observatory "million galaxies" map part of which is shown in Figure 1-7 Shane and Wirtanen analyzed more than one thousand slightly overlapping photographic plates, where each plate was $6° \times 6°$ in size Each plate was divided into $1° \times 1°$ squares for data analysis and each square was further divided into 36 cells 10 arc minutes on a side As it turned out on average there were about 1 5 galaxies per cell, with roughly one third of the cells having no galaxies, one third having one galaxy and one third having two or more galaxies (about 4000 of the million or so cells had 10 or more galaxies) Figure 1-7 is a gray-scale representation of these galaxy counts where white corresponds to no galaxies black to several galaxies.

The raw data behind the original publication (Shane & Wirtanen 1967) were later reprocessed in the "New Reduction of the Lick Catalog of Galaxies" (Seldner Siebers, Groth, & Peebles 1977) This later publication provided the "million galaxy" picture part of which was electronically scanned to make Figure 1-7 Our reproduced image has less resolution than the original image, which has a 16 shade gray-scale corresponding to cell counts of $0, 1, 2, \ldots$ galaxies But even the original has less resolution than the 10 arc minute fineness of the data.

What can we learn by just looking at Figure 1-7? The image seems to reveal some interesting structure. It appears to possess some "signal" as opposed to just "noise." At the risk of using an oxymoron this is structure which is, loosely speaking random or statistical At the same time, it seems obvious that the image is not "random" in

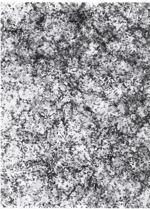


Figure 1.7 Four hundred thousand galaxies—part of the Lick Observatory "million galaxies" survey (Shane & Wirtanen 1967). The galaxy counts are plotted in grey-scale: a black pixel corresponds to many galaxies, a white pixel to no galaxies. This image shows galaxy positions on a section of the celestial sphere within about 50° of the North Galactic Pole. Credit: Seldner et al. (1977)

the colloquial sense, the image does not correspond to the picture one would get by randomly sprinkling flecks of black pepper on a piece of paper.

Remember that this picture shows only angular positions on the celestial sphere. If there is structure here (as discerned by scientific algorithms), what kind of three-dimensional structure does it correspond to?

In 1970, the great Russian scientist Yakov Borisovich Zel'dovich wrote a simple, brilliant paper entitled "Gravitational Instability: An Approximate Theory for Large Density Perturbations." This paper (Zel'dovich 1970) indicated that in an expanding universe, small perturbations from an initial uniform matter distribution would collapse asymmetrically, resulting in flattened "pancakes" where the density of matter was large relative to the mean. Gravitation operates on all scales equally (we say that gravity is "scale-free"), so without added assumptions concerning initial conditions, it is not obvious what size the pancakes would be. Is the flattened structure in Figure 1-4 such an object? Is there a flattening in smaller structures such as clouds of galaxies? Does this pancake instability operate on scales larger than the local supercluster? In the hierarchy of clustering, do smaller structures containing only 10, 100, or 1000 galaxies fit into the pancaking scheme on a scale containing 10,000 galaxies? The answer to all these questions is a qualified "yes" (qualified because the answer really is not very clear). It is tempting to look at Figure 1-7 and perceive a "filamentary" sort of substructure that runs throughout, as might result from large-scale pancaking.

The most satisfying result of this picture comes from statistical measures which have no obvious visual interpretation. We consider the statistical correlations (Peebles 1993) of the galaxy angular positions shown in Figure 1-7. Pick a galaxy from the picture. Then see how many galaxies are 0.5° away from it. Compare this with how many galaxies we would get if the data were from a totally random catalog (e.g. black pepper on paper). There will be either an excess (positive correlation) or scarcity (negative correlation) compared to the random catalog Do this procedure for each galaxy, and for various angles e.g 0 6° 0 7° etc., and then we have the 2 point correlations as a function of angular separation

Here is the important question Do these angular correlations appear to be a projection of full three-dimensional data which are a realization of a homogeneous statistical process? That deserves a translation With our concept of homogeneity we are asking if a chunk of the universe in one place measuring (e g) 300 million light-years on a side is statistically similar to another same-sized chunk in another location To put it another way (still rather cryptically), is the concept of a unique mean density of the universe well-defined? And does the clustering hierarchy finally end so that there are such things as "fair samples" of the universe?

The answer is almost certainly yes. We will stop there only pointing out that the mathematical deduction involves analyzing the angular correlations of various galaxy catalogs to various depths (where bright catalogs are shallow or nearby and dim catalogs are deep) relative to the million galaxy correlations. After comparing the correlations of the catalogs of different depths we see that the data are consistent with a statistically homogeneous universe

In the 1980s a new improved large scale survey was carried out by a team of Oxford astronomers using the UK Schmidt Telescope in Australia. This team used the Cambridge Automated Plate Measuring system to tabulate the angular coordinates and magnitudes of galaxies appearing on 185 photographic plates and produce what is known as the APM survey [13] The survey region was a large field (several thousand square degrees, roughly 50° × 100°) near the South Galactic Pole Galaxies down to a 20 5 magnitude were included so the APM catalog is about three times deeper than

---

[13] For a technical reference including a nice montage of the photographic plates, see Maddox et al (1990)

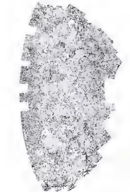


Figure 1 8 Two million galaxies from the Automatic Plate Measuring survey Galaxy counts on the celestial sphere are plotted in grey scale, as in Figure 1 7 This equal area projection covers a contiguous area of 4300 square degrees surrounding the Southern Galactic Pole The APM and Lick survey maps show a frothy texture in the distribution of galaxies due to gravitational clustering Credit Maddox et al (1990), Astrophysics Department, Oxford University the Shane and Wirtanen catalog (which included galaxies down to magnitude 18.5). Figure 1-8 shows a grayscale rendering of the catalog similar to the map of Figure 1-7. There is more information in the APM map, and it is harder to see the frothy signal that seemed to occur in the "million galaxies" map. Note that if we take subsamples of varying depths of the APM catalog, we find angular correlations corresponding to a projection of a three-dimensional homogeneous statistical process, as described above in the case of the Lick Observatory survey. Most importantly, this newer map, culled and analyzed independently, agrees statistically with the Lick survey and with other shallower surveys. This cross checking is crucial and it helps to demonstrate that the large-scale structure problem is being appropriately addressed.

In the late 1970s and early 1980s, the nature of the full three-dimensional structure of the galaxy distribution began to be unveiled. To make a three-dimensional map, we need the redshift of each galaxy in addition to its angular position. Redshift surveys containing hundreds of galaxies during this time seemed to indicate structures—voids (regions containing few galaxies) and chains (narrow regions containing many galaxies)—spanning surprisingly large scales. Could the homogeneous big bang model with gravitational instability really produce structures which span 100 million light-years? While not heretical, such notions of large structures were doubtful at best.

All this changed in 1986 with the publication of "A Slice of the Universe" by de Lapparent, Geller, and Huchra, as part of the CfA (Harvard Center for Astrophysics) survey. Since redshifts are difficult to measure, their survey strategy was to observe galaxies in a thin 6 degrees of declination but to span a large 135 degrees in right ascension. They limited the survey to galaxies brighter than magnitude 15.5, so their survey would not be nearly as deep as the Lick or APM projects. It did not have to be—it was still plenty of work to catalog a few thousand galaxy redshifts and they ended up choosing several "slices" to survey. Their famous first slice is plotted in Figure 1-9. The distance (obtained from the redshift by Hubble's law) is plotted radially. Note that a redshift of 5000 km/s corresponds to a distance of about 230 million light-years (and therefore 10000 km/s corresponds to 460 million light-years). The right ascension is plotted across the top. The six degrees of declination is thin enough that we consider it to be of negligible thickness, and the "slice" is simply plotted on the paper with each dot corresponding to one galaxy. There are 1057 galaxies shown. The map starts to thin out at about 10000 km/s, since at that distance, many galaxies will not appear bright enough to be in the survey.

If the universe contained pancakes or frothy, bubbly structures containing an excess of galaxies, then what would a slice through these types of structures look like? In physics, we can sometimes let the data speak for themselves. Many folks found the data shown in Figure 1-9 to be interesting or even disturbing. If the universe is really homogeneous, then why does that map not look homogeneous? That map is not yet a "fair sample" of the universe. We need to look at a bigger sample, but bigger samples are not easy to come by. However, the sample required cannot be arbitrarily large. If a survey is two or three times bigger than the CfA slice but still seems completely inhomogeneous, then there will be serious theoretical inconsistencies in the overall big bang picture.

Before we move on, let us note a couple of geographical (or rather cosmographical) highlights of these maps. The local supercluster shown back in Figure 1-4 can be partially seen in Figure 1-9. Right next to the apex of the slice, spanning less than a centimeter on the diagram, there is a small chain of galaxies which is part of our local supercluster. Back in Figure 1-5 we saw two highly resolved galaxies which are are part of the Coma cluster. Many of the Coma cluster galaxies are prominently displayed in Figure 1-9. At recession velocity 7000 km/s and right ascension 12.5 hours (right in the upper torso of the humanoid figure), that crowded blob of galaxies

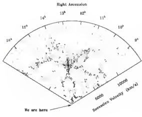


Figure 1.9 A slice of the universe. Each of the 1057 dots represents a galaxy from the CfA redshift survey (de Lapparent et al. 1986). The Milky Way galaxy is not shown—it would be represented by a dot at the apex of the slice. The recession velocity of each galaxy is proportional to its distance from us via Hubble's law. To get an idea of the distances involved, note that a recession velocity of 10 000 km/s corresponds to a distance of about 500 million light-years. The galaxies in this survey are located in a slice which is only 6 degrees in thickness but which carves out a 117 degree swath through space. Dramatic Large-scale Structure is revealed. Data provided in electronic format by Margaret Geller.

is the Coma cluster. Looking at the Million Galaxies map of Figure 1-7, we can pinpoint the Coma cluster if we look carefully. Go to the center of the figure, then down and right by a centimeter or so. That small blob is the Coma cluster, sitting at the North Galactic Pole.

All of the redshifts which have accumulated over the years are a crucial tool in the analysis of the large-scale structure problem. But only large, highly regulated surveys can provide the control necessary statistical analyses. By 1960, there were about 600 published redshifts. By 1976, there were about 2700 measured redshifts, provided by various astronomers studying many different regions of interest. The 1990 total was somewhere around 20 thousand.[14]

Only very recently has a larger sample become feasible. With the electronic detector innovations of Shectman and his colleagues in the Los Campanos Redshift Survey team, they have singlehandedly doubled the redshift database, providing about 26,000 redshifts in six slices. The data were made public by Shectman et al. (1996), and we plot one of the slices in Figure 1-10. The style of plot is the same as the CfA slice plot. The recession velocity (labeled cz) is plotted radially out to 45000 km/s, about three times deeper than the CfA slice. Whereas the declination of the CfA slice covered six degrees of declination centered at 29.5°, this slice is only three degrees thick, centered at a declination of $-12°$, so the slices are forty degrees apart and do not overlap (even though their right ascensions do overlap).

The Los Campanos slice of Figure 1-10 shows 4071 galaxies. While not as visually striking as the CfA slice, it appears to show a fair sample, since it contains many repetitions of structures which are smaller than the slice itself. Perhaps the Los Campanos slices did not merit the same media attention as followed the CfA slice

---

[14] These numbers were obtained from Dressler (1994) and Giovanelli and Haynes (1991).

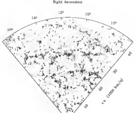


Figure 1.10 A recently obtained slice of the universe. Each of the 4071 dots represents a galaxy from the Los Campanas Redshift Survey (Sheetman et al. 1996). This slice is deeper than the CfA slice and shows a different part of the universe. Note that a recession velocity (obscurely abbreviated as cz in the Figure) of 40 000 km/s indicates a distance of about 1.9 billion light-years, from Hubble's Law.

of 1986 simply because its results were not as controversial. Cosmologists actually expected the results of Figure 1.10, whereas they were a bit shocked by the image of Figure 1.9 when it came out.

Are the redshift data consistent with each other and with the large angular maps? Yes, although the answer is not necessarily trivially established, due to selection effects. Survey galaxies are always scientifically selected for inclusion, yet the surveys are never ideal samples, and comparing surveys can be problematic. Even obtaining a reliable result from one survey can be difficult, as a particular result might be selection-induced or at least selection-biased. In scientific endeavors, the details are often messy, and cosmology is no different than other sciences.

As a relatively simple technical example, we can compare the 2-point spatial correlations. We pick a galaxy from the redshift survey. Then we ask how many galaxies are 10 million light-years, 11 million light-years, etc. from from a galaxy. Is there a positive correlation as compared to a totally random catalog? Continuing in this manner gives us the 2-point correlations as a function of distance. Within uncertainties which must be partly selection induced, the spatial correlations between the two redshift surveys agree, and they correspond to the equivalent angular correlations as obtained from the "million galaxies" map. Interestingly, there does not seem to be anything in the correlation statistics alone that reveals (or would have enabled people to predict, back in the 1970s) the bubbly nature of large-scale structure.

1.6 Statistics and Simulations

The galaxy maps seem to reveal structure. Statistical measures back up this intuitive observation. We believe that the distribution of galaxies over large scales is providing information about all of the matter contained in these large samples of the universe. There is no guarantee of this, but we do need to assume that the galaxies are faithful tracers of the overall matter distribution, which includes dark as well as luminous matter.

Recall that there is physical evidence which supports the big bang model—in particular, that the universe was hotter and smoother in the past. Looking at the redshift maps in Figures 1-9 and 1-10, we see that there are regions, devoid of galaxies, where the matter density is well below the mean, and there are clumpy regions where the density is well above the mean. Hence, we say that today there are large density fluctuations, and we may gather statistics concerning these fluctuations (the 2-point correlations mentioned in the previous section are just one example of the various statistical measures).

Gathering large-scale structure statistics might be rather dry were it not for the high goals of theoretical physics. The COBE satellite data, as well as other modern data, say that the density fluctuations were small in the ancient past, in the radiation era, before there were stars and galaxies. It is natural to attempt to account for the nature of the density fluctuations and their growth over billions of years using the basic tools of physics. Zel'dovich, Peebles and many of their physics colleagues have been dealing with this problem for a while now. This problem includes the thorny issue of galaxy formation. The connection between dynamics (primarily gravitational instability) and the large-scale structure statistics is also part of the problem. Below, we present material related to the dynamics/statistics question.

A simplified mathematical model of the gravitational instability problem will allow us to illustrate some of the features of the large-scale structure problem. Consider a fluid in an expanding universe. The matter content of the universe is represented by the fluid which fills up all of space, and the space itself is expanding according to the precepts of general relativity. As a zeroth order approximation, we say that the universe in this model is exactly homogeneous. The matter density of the universe is the same at all locations and that the fluid velocity is zero everywhere Such an idealization is not ridiculous as it approximately applies to the early universe The mean density (which is the same everywhere) is decreasing with time simply because the universe is expanding and cooling Even though a typical fluid element has no velocity, it appears to be receding from all other points according to the Hubble law since the space itself is expanding

We start off the gravitational instability problem by saying that the initial density is ont perfectly uniform At an early time corresponding to the decoupling era some ten billion years ago we specify that the density is nearly the same everywhere Small fluctuations are imposed as initial conditions The fluctuations are random but not totally random there are spatial correlations in the fluctuations The fluid is considered to be self-gravitating—each fluid element is attracted to any other element by the laws of gravity In this model as we concentrate on one region of the universe, we see that the whole region expands and that the fluid gathers together in a clustering pattern as time goes on When the model reaches the "present" time, there are large fluctuations in the density from place to place, somewhat like the galaxy maps.

Let us show how some computer simulations represent the growth of fluctuations We have limited computer resources so we cannot specify the density everywhere in our simulated universe Instead a "large" region—corresponding to an "observable" patch of this fake universe—millions of light-years in dimension is chosen At perhaps one or two million locations we specify the initial matter density including small fluctuations Then the fluctuations are allowed to grow according to the laws of gravity There are various technical choices which must be made to implement this last task

First we show some results from a simulation in a cubical volume The initial fluctuations were specified at two million locations and at a later time when the fluctuations have grown as the fluid has clustered, a sampling of "galaxies" is taken. Two adjacent slices from a simulation are shown in Figure 1-11. Note that some filamentary structures in one slice are also present in the adjacent slice. If there is a pancake structure in three dimensions, and we take two slices, this is what we would expect.

As a final example, we show some slices obtained from high resolution two-dimensional simulations, and we illustrate statistical differences in an intuitive way.

The matter density is randomly assigned values at one million locations in a slice. Some jargon is helpful in understanding the figures that follow. We speak of a "density fluctuation" or "density contrast" of −0.01 if the density is 1% under the mean. If the density at a particular place is 2% or 200% over the mean, then that location has a density contrast of 0.02 or 2, etc. Initially, all the density fluctuation values are near zero—the fluctuations are small, as they were in the actual early universe. As time time goes by in the simulation and gravity has had its way with the fluid, some matter may pile up into a pancake structure where the density contrast is 1 or 2 (or more). If a region becomes extremely underdense, then at most it would have a negative fluctuation of 100% (a density contrast of −1), corresponding to zero matter density.

Consider the following example which illustrates the "randomness" of the initial conditions. We perform two experiments. In each, we pick 5000 values of the density contrast near zero, distributed according to the bell curve. The results are shown in the two histograms of Figure 1-12. By construction, in each plot, the vast majority of the fluctuation values are near zero (the histograms seem to be centered on zero). If we look closely at both histograms, we see that they are different. But it is intuitively clear that they "have the same statistics," as they both follow the same bell curve.

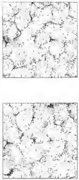


Figure 1-11. Two slices from a computer simulation. Similar to actual redshift surveys, each dot corresponds to a galaxy. Since the slices are close to each other, structures in the top slice carry over into the bottom slice.

(the dotted curve in each figure). In all our simulations, we arrange for the initial fluctuations to be distributed randomly, according the bell curve.

For each set of initial fluctuations, we also specify the 2-point spatial correlations. And as time goes by, the spatial statistics change as matter gets clumped together. Some simulated slices of the universe will help to show the differences in spatial statistics. Figure 1-13 shows a gray-scale plot of the matter distribution in one such simulation, after fluctuations have grown via gravitational instability. Light colors correspond to low density and dark colors to high density, and the numerical fluctuation values are given according to the color bar below the slice. Note there are numerous long pancakes and sparse voids. We did not bother to show a plot corresponding to early times, because the plot would have been nearly all one shade of gray, corresponding to a density contrast close to zero everywhere.

In Figures 1-13 through 1-16 we show four such gray-scale plots of the matter distribution. All of them have initial fluctuations picked according to the bell curve. But the spatial statistics generally differ. In fact, two of the plots have the same spatial statistics. A third plot has statistics close to the these two, but there is a significant mathematical difference. And a fourth plot has statistics which are similar yet clearly different, even to the eye. As a challenge to the reader, try to pick out the four plots so described (the answers are in the next paragraph).

The two statistically identical plots are Figures 1-13 and 1-16. Figure 1-14 looks similar to 1-13 but is in fact significantly different.[15] The statistics of Figure 1-15 are easily seen to be different.[16]

---

[15] The fluctuations in Figure 1-14 have significantly less "large-scale power," in the jargon.

[16] The fluctuations in Figure 1-15 have a different "spectrum" in the large scales, and there are more small-scale fluctuations.

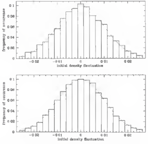


Figure 1-12 Two practically identical distributions of initial density fluctuations. Upon close inspection, these random histograms are different, but they are both realizations of the same bell curve statistical distribution.

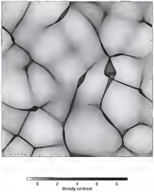


Figure 1.13. The matter distribution in a simulated slice. Light gray corresponds to the mean density (density contrast 0). Higher density regions are darker. These structures evolved from small fluctuations in the initial density.

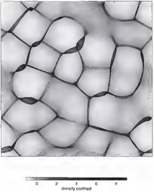


Figure 1-14 The matter distribution in a second simulated slice

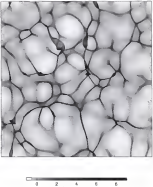


Figure 1.15 The matter distribution in a third simulated slice.

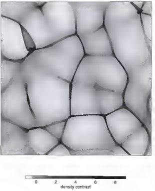


Figure 1-16 The matter distribution in a fourth simulated slice

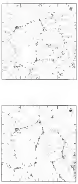


Figure 1-17 Two galaxy maps, each containing about 1000 galaxies from the same matter distribution shown in Figure 1-16 The top map is unbiased with respect to the underlying matter distribution, whereas the bottom map is biased

Taking Figure 1-16 as an example of the dark plus luminous matter distribution, what about the galaxies? We show two galaxy samples in Figure 1-17. There are about 1000 galaxies in each slice. The spatial statistics in the top slice are quite similar to the statistics of Figure 1-16, but the statistical indicators are very "noisy," since the galaxy distribution is discrete while the overall matter distribution is more continuous. Finally, the bottom slice in Figure 1-17 looks similar, but the galaxies are preferentially located in high density regions. The statistics are quite similar between the two slices, but the bottom slice does have measurable statistical differences due to fact that the galaxy distribution is biased with respect to the overall matter distribution.

1.7 Postlude

1.7.1 Final Remarks

It is precisely because the large-scale structure problem is still an epic "work in progress" that we felt compelled to merely give a bit of background and show some pictures which might convey the overall flavor of the problem. Even if it is true that the big bang paradigm is correct and that the gravitational growth of initially small fluctuations provides the correct framework for the large-scale structure problem, there are many loose ends and interesting tidbits to keep scientists busy for many years. It is very important to point out that modern data do support both of the above suppositions. Physicists strive for a complete theories, but concerning the large-scale structure problem we "only" have a generally successful theoretical framework, although this is quite impressive.

We do not wish to give the impression that cosmological work is therefore vague. The experimental observations are as precise as possible. They are just like other measurements in physics, except that there are no repeat experiments (there is just the universe) and the sources of data are so far away. The theoretical work tends to be quite particular. Based on the data, is the galactic halo mostly made up of baryons, or must some of it be non-baryonic? Large clusters of galaxies give off x-ray emissions indicative of hot material; what are the details? The spatial statistics of galaxies are consistent with gravitational instability, but do they distinguish between various theories within the overall framework? Do the statistics show the signs of a bias relative to the overall matter distribution? What is the rate of the current expansion of the universe? The list of questions could get very long.

1.7.2 Bibliographic Note

The best overall reference on cosmology and large-scale structure is arguably "Principles of Physical Cosmology" by Peebles (1993). This is a modern technical reference but is also designed to provide an introduction for physics students. Peebles' approach is conservative, and he makes explicit note at all points where the discussion becomes at all speculative. Though perhaps slightly dated, "The Structure and Evolution of the Universe" by Zel'dovich and Novikov (1975, English edition 1983) is a real tour de force which includes the basics as well as material which was somewhat speculative at the time.

For the layperson, there is Stephen Weinberg's "The First Three Minutes" (1977), which is a challenging read and is widely regarded as scientific writing at its best. Alan Dressler's "Voyage to the Great Attractor" (1994) discusses the large-scale structure problem but also shows an inside view of the sometimes soap-opera-like interactions of scientists. Finally, there are basic books such as "Ancient Light" by Alan Lightman (1991) which popularize the latest developments in cosmology.

Peebles' "The Large-Scale Structure of the Universe" (1980) is a thorough technical reference which also provides extraordinarily detailed bibliographic information

The details and references concerning the history of the large-scale structure problem are thoroughly covered in the two books by Peebles. Also, for historical perspective one may wish to consult "Selected Works of Yakov Borisovich Zel'dovich, Volume II: Particles, Nuclei, and the Universe" (1993). Zel'dovich passed away in December of 1987. For political reasons, he was not widely known outside his home country. He was first allowed to visit the United States in 1980, for a scientific conference. Let us include the following words by astrophysicist Kip Thorne, from a book review in Physics Today:

> The sweep of Zel'dovich's career was breathtaking. . . . My colleagues and I know him as arguably the world's most influential astrophysicist and cosmologist—and also as a creative genius in general relativity and in quantum field theory in curved spacetime. Others revered his contributions to chemical physics, fluid mechanics, the theory of detonations, elementary particle physics, nuclear physics and the design of nuclear weapons. . . . This second volume of Zel'dovich's Selected Works gives some sense of his achievements, breadth, and power. (Thorne 1994, p. 61)

CHAPTER 2
COSMIC STRUCTURE FORMATION

2.1 Big Bang Background

The entire large-scale structure problem of cosmology takes place within the framework of an expanding universe model. The background cosmology is generally relevant to most cosmological questions concerning galaxy formation, large-scale statistics, and the like. However, the minute details are not necessarily always crucial. In the simplest circumstances, there may be background parameters—e.g. the the cosmic expansion rate and the deceleration of the expansion—which are passed down through the large-scale structure calculations. The dependence of cosmic statistical measures on these parameters is sometimes strong, but sometimes not. At the least we need to clarify our notation concerning the background cosmology for use in subsequent calculations.

The material presented in this chapter is part of the standard big bang literature. Concise reports concerning the reliability of the standard model and the status of classical cosmological tests may be found in Peebles et al. (1991) and Sandage (1982), respectively. We include basic elements of the background model below. In the remainder of the chapter, we illustrate the basic physics of structure formation, introduce some of the statistical measures used to describe large-scale cosmic structure, and show the connection between dynamics and the statistics of the large-scale matter distribution. This connection is made through perturbation theory applied to the fluid equations in an expanding background.

The background spacetime is defined by the line element. Using the general theory of relativity we write the invariant interval between two points as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{2.1}$$

where the metric tensor $g_{\mu\nu}$ contains information related to the overall matter and energy content of the universe. The spacetime coordinates x^μ can be somewhat arbitrary. Noting that the universe is expanding and appears to be homogeneous and isotropic, we simply adopt the cosmological principle which formalizes these basic assumptions. For the time coordinate x^0 we use a universal cosmic time t measured by any and all observers "at rest" and thus following the Hubble expansion. Using time-orthogonal coordinates, we write the spatial part of the line element in a manifestly homogeneous and isotropic form to get[1]

$$ds^2 = dt^2 - a^2(t)\left[\frac{dx^2}{1-kx^2} + x^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right] \tag{2.2}$$

This is known as the Robertson-Walker metric and is reviewed by Kolb and Turner (1990) and Peebles (1993). The proofs of the generality of the metric and of homogeneity and isotropy are provided in Weinberg (1972), Hawking and Ellis (1973) and Misner, Thorne and Wheeler (1973).

The spatial manifold is described by the cosmic expansion factor $a(t)$ multiplying a 3-metric for a space of constant spatial curvature. The sign of k in equation (2.2) is the same as the sign of the spatial curvature. The curvature scalar of the 3-dimensional space is the same at all locations by construction, but it depends on time and is given by $k/a^2(t)$, so from this expression and equation (2.2) we see that the overall geometry has a length scale set by $a(t)$. Any coordinate distance $\mathbf{x}$ (specified using integrals over x, θ, and ϕ for example) is related to physical distance by

$$\mathbf{r} = a(t)\,\mathbf{x} \tag{2.3}$$

---

[1] We set the speed of light c to unity.

for $|kx^2| \ll 1$. An arbitrary observer or "galaxy" at fixed x, θ and ϕ is actually in free fall, as a fixed coordinate position is a geodesic of the metric, and therefore a distance $\mathbf{x}$ as in equation (2.3) is called a comoving coordinate.

The matter and energy content of the background are specified using a perfect fluid model. The homogeneous energy density $\bar{\rho}$ depends only on time. The homogeneous pressure p is assumed to be $p = \nu\rho$, where $\nu \approx 0$ corresponds to nonrelativistic matter and $\nu = \frac{1}{3}$ corresponds to ultra-relativistic particles or radiation. Using the general metric (2.2) in the Einstein equations with a perfect fluid stress-energy tensor leads to two independent equations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{k}{a^2} \tag{2.4}$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}G(\bar{\rho} + 3p) \tag{2.5}$$

Newton's gravitational constant is denoted by G. If a $\Lambda g_{\mu\nu}$ term is included in the Einstein equations, then a $\Lambda/3$ is added to the right hand sides of (2.4) and (2.5). From the stress-energy conservation law, we obtain

$$\frac{d}{dt}(\rho a^3) + p\frac{d}{dt}(a^3) = 0 \tag{2.6}$$

which also follows from (2.4) and (2.5). These equations plus a $p(\bar{\rho})$ equation of state describe the $a(t)$ evolution. Note that if $p = \nu\rho$, then

$$\rho(t) \sim a^{-3(1+\nu)} \tag{2.7}$$

so ρ scales like a^{-3} for pressureless matter and a^{-4} for radiation.

The redshifts of distant objects such as galaxies provide us with observable information related to the scale factor $a(t)$. The redshift z is defined as the fractional change of the observed wavelength λ_0 relative to the wavelength λ_e emitted at the galaxy

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_e} \tag{2.8}$$

The wavelength λ_e is known once the spectral line has been identified, since by the cosmological principle, distant systems have the same physical laws. According to the geodesic equation, the momentum of a free particle is proportional to $1/a(t)$. That is, the wavelength of a photon stretches, along with other cosmological lengths, in proportion to $a(t)$. Therefore, for an observed redshift z,

$$1 + z = \frac{\lambda_0}{\lambda_e} = \frac{a(t_0)}{a(t_e)} \tag{2.9}$$

where here, and below, we use a subscript 0 for our present epoch. It is customary to refer to cz as a recession velocity, in order to make an intuitive correspondence between this cosmological redshift and the usual motion-induced redshift. We can use z as indicator of past epochs. Today corresponds to $z = 0$, the most distant galaxies and quasars have redshifts of $z \approx 4$, and the last scattering of the cosmic background radiation at decoupling occurred at $z \approx 1000$.

The rate of expansion is given by the Hubble parameter

$$H = \frac{\dot{a}}{a} \tag{2.10}$$

At the present epoch this is called the Hubble constant H_0,

$$H_0 = 100\, h \,\mathrm{km\,s^{-1}\,Mpc^{-1}} \approx h/(10^{10}\,\mathrm{years}) \tag{2.11}$$

Throughout our discussion we use the megaparsec (Mpc) unit

$$1\ \mathrm{Mpc} \approx 3.086 \times 10^{19}\,\mathrm{km} \approx 3.26 \times 10^{6}\ \text{light-years}. \tag{2.12}$$

The large uncertainty in the measurable quantity H_0 is relegated to the dimensionless parameter h, with a rather large window $0.4 \lesssim h \lesssim 1$. Sandage (1993) and Tully

(1993) summarize some of the recent H_0 determinations. The entire distance scale question is still at issue, and that is why we often refer to a distance as $5\,h^{-1}$ Mpc rather than using a favored value such as $h \approx 0.7$ stemming from recent measurements.

The density in the $k = 0$ case is called the critical density $\rho_c(t)$, and from the Friedman equation (2.4), we have

$$\rho_c = \frac{3H^2}{8\pi G} \tag{2.13}$$

The dimensionless density parameter $\Omega(t)$ is defined by

$$\Omega = \frac{\rho}{\rho_c} \tag{2.14}$$

The spatial curvature is directly related to the mean density of the universe. Using the definitions of H and Ω in equation (2.4) we get $k = a^2H^2(\Omega - 1)$, which clarifies how the flat geometry at critical density $\Omega = 1$ is the boundary between the negatively and positively curved spaces.

The evolution of a(t) is is also directly related to the mean density from equation (2.4). For $p = \nu\rho$, with a conservative constraint $\nu > -\frac{1}{3}$, integration of equation (2.4) says that there was a singularity sometime in the finite past and that the universe will expand forever if $k \leq 0$ and that the expansion will someday reverse if $k > 0$.

In the flat geometry case, solutions for a(t) are given by $a \sim t^{2/3(1+\nu)}$. In the early radiation dominated era, $a \sim t^{1/2}$. In the current matter dominated era (which began shortly before decoupling), $a \sim t^{2/3}$. In the material that follows, we focus solely on the matter dominated ($p \approx 0$) case which is the relevant background for the growth of fluctuations which we study. Since the matter era dwarfs the radiation era in duration, we can go ahead and use the matter dominated solution to calculate the the current cosmic time $t_0 = \frac{2}{3}H_0$ and write

$$a(t) = a_0 \left(\frac{t}{t_0}\right)^{2/3} \tag{2.15}$$

The $a(t)$ solutions for $k \neq 0$ and $p = 0$ (cf. Weinberg 1972 and Zel'dovich & Novikov 1983) start out (in the limit $t \to 0$) with $a \sim t^{2/3}$ and $\Omega = 1$ but then diverge. We plot three $a(t)$ solutions in Figure 2.1. For convenience we set the current ($t = t_0$) expansion factor $a_0 = 1$ and we choose $h = 0.5$ in the current H_0. The open ($k = -1$) model has $\Omega_0 = 0.1$; its Ω decreases with time. The closed ($k = 1$) model has $\Omega_0 = 2$; its Ω will increase with during expansion. The dimensionless deceleration parameter $q = \ddot{a}a/\dot{a}^2$ is related to the density, using equation (2.5), by $q = \Omega/2$. The current age t_0 is a function of Ω_0 such that t_0 is 11 billion years in the closed model, 13 billion years in the flat model, and 18 billion years in the open model.

The Hubble "$v = Hr$" law is in place once we have $r = a(t)x$, as in equation (2.3), since this gives $\dot{r} = Hr$. But $\dot{r}$ is not measured. The cosmological redshift z is the measurable quantity, and all the Friedman models give, to first order in z,

$$cz = H_0 r \tag{2.16}$$

For completeness, we record the full equation relating a physical distance $r = a_0 x$ to a redshift z (Mattig 1958):

$$H_0 r = \frac{2c}{\Omega_0^2(1+z)} \left\{ \Omega_0 z + (\Omega_0 - 2)\left[(\Omega_0 z + 1)^{1/2} - 1\right] \right\} \tag{2.17}$$

$$\approx c\left[z - z^2(1/2 + \Omega_0) + \dots\right] \quad \text{for } z \ll 1 \tag{2.18}$$

In our observable section of the universe, the global background model appears to be a Friedman model with $0.1 \lesssim \Omega_0 \lesssim 2$. A nonzero cosmological constant could still be considered a possible addition to the model, but opinions vary in the literature.

The present geometry, if not flat, is remarkably close to flat. This can be seen from the Friedman equation. We use equation (2.4) in the form that says $k = a^2 H^2(\Omega - 1)$ is a constant in time, combined with the density evolution which says that $\rho a^{3(1+\nu)}$

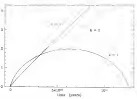


Figure 2.1 The expansion factor $a(t)$ vs t for open ($k = -1$), closed ($k = 1$) and flat ($k = 0$) models. These are matter dominated solutions of the Friedman equation (2.4) and for illustrative purposes, we chose solutions so that $a(t_0) = 1$ corresponds to the present time. At t_0, we fixed H_0 with $h = 0.5$ and we set $\Omega(t_0) = 0.1$ (open) and $\Omega(t_0) = 2$ (closed). For the flat model $\Omega = 1$ for all t and $a \sim t^{2/3}$ with t_0 equal to 13 billion years.

remains constant. So the present Ω_0 at redshift $z = 0$ is related to an earlier Ω at redshift z by

$$\Omega^{-1} - 1 = \frac{(\Omega_0^{-1} - 1)}{(1 + z)^{1+3w}} \tag{2.19}$$

For example, if $\Omega_0 = 0.1$, then at decoupling ($z \approx 1000$) the value was $\Omega = 0.99$, and in the very early universe Ω had to be finely tuned to a value very close to 1. The fine tuning seems like an odd coincidence, and this is referred to as the flatness problem in early universe studies. Incidentally, this indicates that the Einstein–de-Sitter model ($k = 0$, $p = 0$, $\Lambda = 0$) is a good approximation of earlier times and therefore it is used in many calculations.

The inflationary theory, which is a very early universe addendum to the standard big bang model, generally predicts $\Omega_0 = 1$ to high accuracy. A momentary cosmological constant inflated $a(t)$ exponentially, causing the k/a^2 and (non vacuum) p terms to be negligible. Inflation also addresses the causality or horizon problem in the initial conditions of our patch of the universe.

The causal horizon—the sphere within which $ds^2 = 0$ signals are observable—has radius d_{hor} given by

$$d_{hor} = a(t) \int_0^t \frac{c\,dt'}{a(t')} \tag{2.20}$$

and we have $d_{hor} \sim ct$ for the Friedman models (e.g. with radiation or matter). We often refer to Hubble radius $c/H \sim ct$ as the "horizon", and this has the same order of magnitude as d_{hor} for our simple Friedman models. Since $d_{hor} \sim t$ and $a \sim t^{\alpha}$ ($\alpha < 1$), a region of proper size d (which scales like a), lies outside the horizon until such time as $d_{hor} = d$. Photons in the cosmic background radiation separated by the horizon distance at decoupling ($z \approx 1000$) reach us today separated by less than a degree on the sky. The uniformity of the cosmic background radiation indicates a homogeneity which naively would violate a conventional causal origin. This is a major puzzle. In the large-scale structure problem, initial conditions are often simply assumed as given at decoupling, so this issue is put on the back burner. But inflation does appear to deal with it nicely. The causal structure is reworked during the brief accelerated ($\ddot{a} > 0$) inflationary phase which stretches causally related regions to scales well beyond our present horizon. While providing for large-scale homogeneity, the inflationary scenarios also predict Gaussian metric fluctuations whose imprint lasts throughout the subsequent Friedman model evolution up to matter domination.

The current temperature of the universe, as defined by the thermal bath of cosmic background radiation photons, is 2.7 K. To be exact, the COBE FIRAS team has determined $T = 2.726 \pm 0.010\,\mathrm{K}$ (Mather et al. 1993) in an exquisite fit to a Planck spectrum—the most precise measurement of a Planck spectrum in nature turns out to pertain to the relic radiation of the big bang. This background presently contains about 400 photons per cubic centimeter, and the universe has been effectively transparent to these photons since decoupling, at which time the photons decoupled from the matter and continued their adiabatic cooling (from $T \approx 1\,\mathrm{eV} \approx 10^4\,\mathrm{K}$) down to their current 2.7 K. The universal thermal history from the hadron era ($T \approx 100\,\mathrm{MeV}$) or even earlier (inflation, baryogenesis), through nucleosynthesis ($T \approx 1\,\mathrm{MeV}$), continuing through decoupling is reviewed thoroughly by Peebles (1971, 1993), Zel'dovich and Novikov (1983), and Kolb and Turner (1990). Since the particle physics astrophysics interface is so well described in the celebrated volume by Kolb and Turner, we choose to quote their apt summary concerning the thermodynamic background in cosmology:

> For much of its early history, most of the constituents of the Universe were in thermal equilibrium, thereby making an equilibrium description a good approximation. However, there have been a number of very notable departures from thermal equilibrium—neutrino decoupling, decoupling of the background radiation, primordial nucleosynthesis, and on the more speculative side, inflation, baryogenesis, decoupling of relic WIMPs, etc. As we have previously emphasized, if not for such departures from thermal equilibrium, the present state of the Universe would be completely specified by the present temperature. The departures from equilibrium have led to important relics—the light elements, the neutrino backgrounds, a net baryon number, relic WIMPs

(weakly interacting massive particles) relic cosmologists and so on (Kolb & Turner 1990 p 115)

We next present thermodynamic material which helps to set the context for the gravitational instability problem The stage is set for structure formation during the radiation dominated period and so we begin there At this point fluctuations in the metric (that is to say, perturbations in the gravitational potential) are present but not free to grow After the transitional period of matter-radiation equality some of the fluctuations in the matter distribution—weakly interacting nonbaryonic dark matter in particular—are free to amplify according to gravitational instability Prior to decoupling pressure forces supplied by the photon radiation are sufficient to counteract the effect of gravity on the baryons but after decoupling the baryons too can participate in the instability on cosmologically interesting scales We introduce the physics of gravitational instability in the next section

Approximately three minutes after the big bang, the radiation dominated era is in full effect The last important period ($t \approx 10^{-2}$–10^{2} s or $T \approx 10$–10^{-1} MeV) prior to this time gave us the annihilation of the $e^{\pm}$ species (leaving enough electrons to maintain neutrality with respect to the remaining protons) and the nuclear reactions which were able to construct Helium nuclei ($\approx 25\%$ abundance) and small amounts of ^{2}D, ^{3}He and ^{7}Li from the protons and neutrons leaving $\approx 75\%$ Hydrogen nuclei (protons)

After nucleosynthesis the equilibrium soup contains photons (γ), neutrinos (ν) electrons (e^{-}) and baryons (B) We define T to be the photon temperature The (massless) neutrinos are no longer really in equilibrium having decoupled at $T \approx$ 1 MeV, prior to which their weak interactions had kept them in equilibrium Also present is the nonbaryonic dark matter which we believe to have been in equilibrium with the photons at some point in their history but which similarly to the neutrinos decouple from the photons and ordinary matter early on The equilibrium condition is not true in the case of axions, which do not freeze out in the thermodynamic sense but fall into the soup as a result of an arcane phase transition (cf. Kolb & Turner 1990). The energy density ρ of the universe during this era comes from the γ and ν radiation

$$\rho = \frac{\pi^2}{30} g T^4 \qquad (2.21)$$

where g is an effective count of the degrees of freedom due to the number of species and their spin. The pressure is $p = \frac{1}{3}\rho$ due to photons. The second law of thermodynamics applied to a comoving volume element implies that the entropy density (contributed primarily by the relativistic species) is

$$s = \frac{\rho + p}{T} = \frac{2\pi^2}{45} g_S T^3 \qquad (2.22)$$

during equilibrium.[2] Since the entropy sa^3 is constant, we have

$$T \sim a^{-1} \qquad (2.23)$$

for such periods. The entropy contributions from photons and neutrinos have been separately conserved since neutrino decoupling.

The present value of the critical density ρ_c is

$$\rho_c(t_0) \approx 1.88 \times 10^{-29}\, h^2\, \mathrm{g\,cm^{-3}} \qquad (2.24)$$

Using equation (2.21) with $T_0 \approx 2.7\,\mathrm{K}$ gives $\Omega_{rad} \approx 2.6 \times 10^{-5} h^{-2}$ as a contribution to the overall density, but we know the density today is $\Omega_0 \approx 1$ (even with no

---

[2] The ρ and s degeneracy factors are

$$g = 2 + 2\,\frac{7}{8}\,(\frac{4}{11})^{4/3}\,3$$
$$g_S = 2 + 2\,\frac{7}{8}\,\frac{4}{11}\,3$$

nonbaryonic dark matter, we have $\Omega_0 \approx 0.1$). The universe is currently matter dominated. Direct observations of starlight from the luminous parts of galaxies give (from galaxy counts, the galaxy luminosity distribution, and the mass to light ratio) $\Omega_{\rm lum} \approx 0.01$. Such an estimate uses a mass to light ratio of $\approx 10h$ in solar units ($M_\odot/L_\odot$), but dynamical estimates for larger systems, from galaxy halos to clusters to large-scale dynamics, give larger mass to light ratios (Peebles 1993). That dark matter exists has long been known, but the surprise is perhaps that the dark matter seems to comprise much more of the universe than luminous matter. The value of Ω_0 is not well determined, but estimates tend to give $\Omega_0 \gtrsim 0.1$. Note that big bang nucleosynthesis calculations (which fare well with observational light element abundances) place an upper bound on the baryon density. Also, simple baryon only models have severe problems related to structure formation. Finally, carrying the least weight in a strict argument perhaps, inflationary models generally predict that $\Omega_0 = 1$ to high precision. Hence some, possibly most, of the dark matter must be nonbaryonic.

Gravitational instability is free to operate in the matter dominated era, as we illustrate in §2.2.2. We can estimate the era of matter-radiation equality. If we simply equate the radiation density $\frac{1}{2} g T_0^4 (1+z)^4$ to the matter density $\rho_c(t_0)\Omega_0(1+z)^3$ we find the era of equality to be $z = 4 \times 10^4 \Omega_0 h^2$. This transition occurred shortly before decoupling ($z \approx 1000$).

We end this section with an attempt to add a little historical perspective. The historical development of physical cosmology and the large-scale structure problem is well covered by Peebles (1980, 1993). This history is not always well appreciated, as noted by Fred Hoyle (1993) in an assessment of modern cosmology and its history: "There is a tendency nowadays to think that cosmology only began as a serious study in 1965, whereas in fact rather little that is really new has happened since 1965" (Hoyle

1993 p 693) George Gamow (1948a b) had worked out essential aspects of the theory (e.g. background radiation and a globular cluster Jeans mass) decades before Penzias and Wilson (1965) published their cosmic background radiation discovery On the other hand there really was a shift towards a more intense attack on cosmology in general at this time as physicists became more inclined to use a more rigorous approach to this formerly somewhat tentative science This sentiment is reflected by Zel'dovich

> There is every reason to assume that the initial stage in the evolution of the Universe corresponded exactly to the solution offered by Friedmann We assume that there was a moment of infinite density at $t = 0$ which is a singular solution The existence of the infinite density can be regarded as essentially a question of belief If so *I do believe* and shall attempt to draw all possible conclusions from this belief comparing these conclusions with the facts observed until some irresistible contradiction dissuades me So far I do not know of any such contradiction (Zel'dovich 1965 p 313)

2.2 Cosmic Structure

2.2.1 Basic Ideas

We live in a section of the universe which appears homogeneous and isotropic, when averaged with a large enough smoothing filter Adopting the cosmological principle allows us to use a big bang background model for the overall evolution of the universe and we consider fluctuations with respect to this background

In regions smaller than an individual galaxy, we know that the matter distribution is highly structured molecules planets stars, Even if we look at large-scale maps of the galaxy distribution as shown in Chapter 1 the distribution is still quite lumpy The universe was quite uniform in the radiation dominated era There were no galaxies stars or heavy elements Small amplitude fluctuations did exist, perhaps as a relic of an inflationary era These were real fluctuations with a statistical signal we presume [3] and not simply fluctuations associated with the discreteness of matter (such "minimal" fluctuations do have their own signal, too small to detect). We adopt the standard scenario for structure formation whereby the small fluctuations grew via gravitation into the observed large-scale structure. By focusing on large scales, gravitational instability is the key ingredient of any model. The details of hydrodynamics and electromagnetism can never be ignored, but by not considering galaxy formation per se, cluster or halo composition, etc., we effectively side-step such issues (in the particular areas which we study, not in the large-scale structure problem in general).

We now begin to describe deviations from the perfectly homogeneous background and thus begin our focus on the large-scale structure problem. To get a sense of the numbers, we use rough orders of magnitude for presently observed values of the mass density, which we denote by ρ. On Earth, we have $\rho \approx 1\,\mathrm{g\,cm^{-3}}$, while for the Solar System, $\rho \approx 10^{-10}\,\mathrm{g\,cm^{-3}}$. For a galaxy, we make a crude estimate[4] that $M \approx 10^{12} M_{\odot}$ and $d \approx 0.01\,\mathrm{Mpc}$ to arrive at $\rho \approx 10^{-24}\,\mathrm{g\,cm^{-3}}$. The mean cosmic density is $\rho \approx 10^{-29} - 10^{-30}\,\mathrm{g\,cm^{-3}}$. On cosmic scales it is convenient to use the dimensionless density contrast δ,

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \qquad (2.25)$$

At the present epoch, large structures have fluctuations ranging from $\delta \approx 1$ to $\delta \approx 10^{6}$, as we indicate in Table 2.1.

As a consistency check, we can estimate that structure formation (galaxies, clusters, ...) was, in fact, a relatively recent occurrence—e.g., that such structures didn't form in the radiation era or near decoupling. Since ρa^{3} is constant in the matter

---

[3] Since the COBE detection of cosmic background radiation fluctuations (Smoot et al. 1992) we believe we have direct evidence of such a signal.

[4] The mass of the sun is $M_{\odot} \approx 2 \times 10^{33}\,\mathrm{g}$.

Table 2 1 Density fluctuations for large objects

Density Contrast	Object	Size [h^{-1} Mpc]
10^6	Galaxy	0 01
10^3	Galaxy Cluster	1
few	Galaxy Supercluster	100
0	Fair Sample	1000

dominated era, we can write for earlier times that $\rho = \rho_0(1+z)^3$ Since we have $\delta \approx 10^6$ 10^3 and 1 for galaxies clusters and superclusters, we can simply trace ρ back to higher z and we get redshifts of $z \approx 100$ 10, and 1 for the corresponding mean densities. If such structures had separated out earlier (at higher z) they would clearly have higher densities presently A conservative detailed estimate of the timetable is given by Peebles (1993 §25) and we may take $z \approx 5$ as the galaxy formation era, and $z \approx 1-3$ as the era for larger structures

The apparent motion of a galaxy at $\mathbf{r} = a\mathbf{x}$ is divided into two pieces as in

$$\dot{\mathbf{r}} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}} = H\mathbf{r} + \mathbf{v} \tag{2 26}$$

The Hubble flow $H\mathbf{r}$ corresponds to the geodesic solution of a point at rest in the expanding background The velocity $\mathbf{v}$ relative to the mean rest frame is known historically as the peculiar velocity The peculiar motions within our local group and our infall toward the Virgo Cluster are nearby examples of measurable $\mathbf{v}$ values. The "Fingers of God" effect easily seen in the Coma cluster in the Slice of the Universe Figure 1-9 results from the fact that the radial components of $\mathbf{v}$ for galaxies get included in the redshift measurements and distort the redshift map relative to a pure galaxy position map In principle the velocity field $\mathbf{v}$ is detectable in redshift surveys—once the distance (e g from the Tully Fisher relation) has been subtracted out of the redshift—since the radial component of $\mathbf{v}$ is sufficient to extract the entire

$\mathbf{v}$ if the velocity field is a gradient (which it is if gravitational instability is operating on the large). Current $\mathbf{v}$ field results are preliminary, but the technique will have great power since the entire matter distribution (whether luminous or not) is the source of the flow (Bertschinger & Dekel 1989). Equation (2.26) also holds good in our analysis of the growth of small fluctuations in a fluid model, as $\mathbf{v}$ will be the velocity field with respect to the expanding background.

In the fluid model, our notation will resemble that of conventional fluid mechanics. The density field $\rho(\mathbf{x},t)$ depends on space and time, and we use comoving coordinates $\mathbf{x}$, so that $\mathbf{x}$ would be referred to as a spatial coordinate (as opposed to a material coordinate).[5] We consider variations from the background $\bar{\rho}(t)$, and the subsequent dynamical equations are more conveniently expressed using $\delta(\mathbf{x},t)$, the density contrast. In the equations of motion for the Newtonian approximation, introduced below, the gravitational potential φ is connected to δ via the Poisson equation, and $\mathbf{v}(\mathbf{x},t)$, $\delta(\mathbf{x},t)$, and $\varphi(\mathbf{x},t)$ are related using the continuity and Euler equations. The Eulerian approach is more natural in the sense that measurements (e.g. redshift surveys) supply information related to the fields $\delta(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$ at the current epoch. Later, particularly in Chapter 3, we have occasion to use the Lagrangian method, where the separation between uniform expansion and peculiar motion is used directly, in the form

$$r(\mathbf{q},t) = a(t)\left[\mathbf{q} + \Psi(\mathbf{q},t)\right] \tag{2.27}$$

---

[5] The spatial coordinate formulation of fluid mechanics is known as "Eulerian", while the equivalent material coordinate formalism is known as "Lagrangian". This terminology is odd, considering Euler developed both methods prior to Lagrange and that Lagrange's contributions were primarily to the spatial coordinate method. This is clarified by Truesdell (1954). In spirit, the names are appropriate (though not accurate), and we keep with the tradition, as we have occasion to use both methods.

where material coordinates **q** are specified when fluctuations are small, and the gravitational instability is described by the function $\Psi(\mathbf{q}, t)$ which supplies the deviations from the Hubble flow and which leads to structure.

Corresponding to any comoving length scale λ is a wavevector **k**, $k = 2\pi/\lambda$, in Fourier transform space. The density contrast is considered as the sum of plane wave[6] modes $\delta(\mathbf{k}, t)$:

$$\delta(\mathbf{x}, t) = \int \frac{d^3k}{(2\pi)^3} \delta(\mathbf{k}, t) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.28}$$

The Fourier decomposition proves useful in characterizing the initial density fluctuations and their evolution. In addition, the current matter distribution, as traced by galaxies, can be usefully described by statistics in both the **x** or **k** domains. It is often convenient to normalize the expansion factor $a(t_0)$ to unity, so the comoving and physical scales coincide at the present epoch. Tracing a length scale backwards is possible, since the physical scale is $a(t)\lambda$.

2.2.2 Jeans Instability

The Jeans (1902, 1928) theory of gravitational instability is fundamental for a basic understanding of the onset of formation. The Jeans instability is covered particularly well in two references: Peebles (1980), who also covers its history as applied to cosmology, and Zel'dovich and Novikov (1983, Chapter 10), who provide details for both a static and an expanding medium. We introduce some of the physics below.

The essence of the idea is most easily illustrated following Jeans' approach for a collisional fluid in a static background—he intended to apply the results to planet and star formation within a spiral nebula. The gravitational equations are certainly scale-free, and so the concepts transfer freely to our problem, though there are two key reasons why the results do not directly apply to cosmology: the universal expansion is neglected and the methods are purely Newtonian. The first problem is crucial and must be corrected for. The second is not really a setback, as the Newtonian approximation which we ultimately use in our cosmological analysis is an appropriate limit (subhorizon and slow motion) of the full general relativistic analysis.

---

[6] The $e^{i\mathbf{k}\cdot\mathbf{x}}$ mode expansion is strictly valid only when the physical scale of the modes is much less than the radius of curvature of the background geometry. But this is the case now (even if we are not in an Einstein–de-Sitter model) and even more so in the past. In the Newtonian regime in which we operate, our wavelengths are much smaller than the horizon, which addresses other caveats concerning such a mode expansion (Kolb & Turner 1990).

The fluid equations (continuity, Euler, and Poisson) plus a thermodynamic statement (dissipationless flow—S is the entropy per unit mass) are

$$\frac{\partial\rho}{\partial t}+\nabla\cdot\rho\mathbf{u}=0 \tag{2.29}$$

$$\frac{\partial\mathbf{u}}{\partial t}+\mathbf{u}\cdot\nabla\mathbf{u}+\frac{1}{\rho}\nabla p+\nabla\Phi=0 \tag{2.30}$$

$$\nabla^2\Phi=4\pi G\rho \tag{2.31}$$

$$\frac{\partial S}{\partial t}+\mathbf{u}\cdot\nabla S=0 \tag{2.32}$$

First take a constant background solution (density $\rho=\rho_b$, velocity $\mathbf{u}=0$, etc) and add perturbations, whose solutions we may seek as a sum of plane wave modes as in equation (2.28), yet we write one mode only for simplicity:

$$\rho(\mathbf{x},t)=\rho_b[1+\delta(t)e^{i\mathbf{k}\cdot\mathbf{x}}] \tag{2.33}$$

$$\mathbf{u}(\mathbf{x},t)=0+\mathbf{v}(t)e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.34}$$

$$\Phi(\mathbf{x},t)=\Phi_b+\phi(t)e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.35}$$

$$S(\mathbf{x},t)=S_b+s(t)e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.36}$$

$$p(\mathbf{x},t)=p_b+v_s^2(\rho-\rho_b)+\frac{\partial p}{\partial S}(S-S_b) \tag{2.37}$$

$$v_s^2=\frac{\partial p}{\partial\rho} \tag{2.38}$$

Placing these perturbed expressions back into the fluid equations (2 29–2 32) while assuming the background solutions satisfy the equations identically and the second order terms can be neglected [7] we arrive at the following equations for one $\mathbf{k}$ mode,

$$\delta + i\mathbf{k}\cdot\mathbf{v} = 0 \tag{2 39}$$

$$\mathbf{v} + i\mathbf{k}\phi + i\mathbf{k}v_s^2\delta + i\mathbf{k}\frac{1}{\rho_b}\frac{\partial p}{\partial S}s = 0 \tag{2 40}$$

$$-k^2\phi = 4\pi G\rho_b\delta \tag{2 41}$$

$$s = 0 \tag{2 42}$$

The background is static, and ρ_b, v_s^2 and $\mathbf{k}$ are time-independent in equations (2 39–2 42) [8] and we find that an exponential $e^{\omega t}$ time dependence holds good for the perturbation solutions

We focus on the possibly unstable adiabatic solutions to equations (2.39–2 42) That is $\omega \neq 0$, and therefore equation (2 42) implies that $s = 0$ We write $\delta(t) = \delta_0 e^{\omega t}$ $\mathbf{v}(t) = \mathbf{v}_0 e^{\omega t}$ etc and split $\mathbf{v}_0$ into one longitudinal (curl-free) and two transverse (rotational) parts

$$\mathbf{v}_0 = (\mathbf{k}\cdot\mathbf{v}_0)\mathbf{k} + \mathbf{k}\times(\mathbf{v}_0\times\mathbf{k}) \tag{2 43}$$

Only the curl-free mode is unstable The rotational modes are frozen in (they are $\omega = 0$ solutions) Therefore we write $\mathbf{v}_0 = v_0\mathbf{k}$. The perturbation solution must satisfy equations (2 39–2 41), which we rewrite as

$$\omega\delta_0 + ikv_0 = 0 \tag{2 44}$$

$$\omega v_0 + ik\phi_0 + ikv_s^2\delta_0 = 0 \tag{2 45}$$

$$k^2\phi_0 + 4\pi G\rho_b\delta_0 = 0 \tag{2 46}$$

---

[7] The background "solution" is in fact a solution except that we need to take $\nabla\Phi_b = 0$ which violates the the Poisson equation This is humorously known as the "Jeans swindle"; but there is no such violation in an analysis with an expanding background

[8] In the analogous mode equations for an expanding fluid these quantities would depend on time due to the $a(t)$ factor

and a non trivial solution for δ_0, $\mathbf{v}_0$ and ϕ_0 exists only if the corresponding determinant is zero, which gives the following dispersion relation

$$\omega = \pm\sqrt{4\pi G\rho_b - v_s^2 k^2} \tag{2.47}$$

It is clear from equation (2.47) that there are two types of solutions, as the length scale of the perturbation may be greater than or less than the critical Jeans length

$$\lambda_J = \frac{2\pi}{k_J} = v_s\sqrt{\frac{\pi}{G\rho_b}} \tag{2.48}$$

The unstable solution occurs for small k (scales larger than λ_J) when ω is real and positive, and we may write out the solutions from equations (2.44–2.46)

$$\delta(\mathbf{x}, t) = \delta_0 e^{i\mathbf{k}\cdot\mathbf{x}} e^{\omega t} \tag{2.49}$$

$$\mathbf{v}(\mathbf{x}, t) = i\frac{\mathbf{k}}{k^2}\delta_0 (4\pi G\rho_b)^{1/2}\left[1 - (k/k_J)^2\right]^{1/2} e^{i\mathbf{k}\cdot\mathbf{x}} e^{\omega t} \tag{2.50}$$

$$\phi(\mathbf{x}, t) = -\delta_0 \frac{4\pi G\rho_b}{k^2} e^{i\mathbf{k}\cdot\mathbf{x}} e^{\omega t} \tag{2.51}$$

In particular, the δ and $\mathbf{v}$ solutions are spatially out of phase by $\pi/2$, so, for example, a velocity maximum coincides with a zero in δ, as one would expect. Solutions with $\omega < 0$ are, in fact, solutions, but they are related to particulars of the initial conditions and nonetheless get swamped by the unstable solutions as t increases. Clearly the solutions (2.49–2.51) are physical (as opposed to mathematical artifacts) as it is energetically favorable for the unstable solutions to grow. Such an instability increases without bound (there is nothing in the linearized equations to stop it), but actual growth has stopped when an internal potential/kinetic energy balance is reached in the separated object, and a bound system such as a star or galaxy is made. And when there is dissipation, the collapse to a bound object is even deeper (the gravitational potential is more negative, but the kinetic energy is larger, as per the virial theorem). Large k modes ($\lambda < \lambda_J$) propagate like a sound wave (exactly in the limit $k \gg k_J$) with frequency $v_s k[1-(k_J/k)^2]^{1/2}$ There are two waves in the $\pm\mathbf{k}$ directions and no instabilities

The qualitative features shown above for a fluid with pressure transfer well to the cosmological scenario as related to baryons before decoupling when pressure is supplied by the hot photon gas The physical situation is more accurately described by a collisionless fluid in the matter dominated era for weakly interacting dark matter (or for baryons after decoupling) Such a model uses a smooth gravitational potential Φ and density ρ related by the Poisson equation describing a sea of noninteracting particles of mass m The phase space distribution function of the particles is denoted by $F(\mathbf{x}, \mathbf{u}, t)$ and with a phase space trajectory determined by

$$\dot{\mathbf{x}} = \mathbf{u} \quad \dot{\mathbf{u}} = -\nabla\Phi \tag{2 52}$$

the dynamics are governed by the Vlasov equation

$$\frac{\partial F}{\partial t} + \nabla \cdot (F\mathbf{u}) + \nabla_{\mathbf{u}} \cdot (F\dot{\mathbf{u}}) = 0 \tag{2 53}$$

This model is described by Peebles (1980 §9) and Zel'dovich and Novikov (1983 §10 7) and we show the main ideas which lie at the foundations of the gravitational instability problem What we present below resembles the treatment of the simplest macroscopic small amplitude plasma disturbances (Ichimaru 1986, Goldston & Rutherford 1995) since the plasma discussion also begins with the Vlasov equation and leads to similar dispersion relations—with crucial differences since the gravitational coupling and the inertial mass are the same m and gravity is always attractive in obvious contrast to the situation in a plasma

Using similar notation to that in the above collisional analysis, we perturb about a stationary background solution [9] Denoting the perturbations by f δ $\mathbf{v}$, and ϕ we have

---

[9] The same "swindle" involving the Poisson equation is used, but again this problem does not persist in the case of expanding background

$$f(\mathbf{x}, \mathbf{v}, t) = F_b(v^2) + f(\mathbf{x}, \mathbf{v}, t) \qquad (2.54)$$

and the linearized version of equation (2.53) is

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f - \nabla \phi \cdot \nabla_{\mathbf{v}} F_b = 0 \qquad (2.55)$$

with relations such as

$$\mathbf{v} = -\nabla \phi \qquad (2.56)$$

$$\nabla^2 \phi = 4\pi G \rho_b \delta \qquad (2.57)$$

$$\rho_b \delta = m \int f \, d\mathbf{v} \qquad (2.58)$$

As before, we seek plane wave solutions of the form $e^{ikx}e^{\omega t}$ for the perturbations (where we pick $\mathbf{k} = k\mathbf{x}$ without loss of generality). Placing perturbations of this form ($f = f_0(\mathbf{v})e^{ikx+\omega t}$, $\delta = \delta_0 e^{ikx+\omega t}$ etc.) in equation (2.55) gives a formal solution

$$f_0(\mathbf{v}) = -ik \left[\frac{4\pi G \rho_b \delta_0}{k^2}\right] \frac{2v_x}{\omega + ikv_x} \frac{dF_b}{dv^2} \qquad (2.59)$$

which when put into equation (2.58) gives a somewhat complicated dispersion relation for $\omega(k)$

$$-\frac{8\pi G m}{k^2} ik \int \frac{v_x}{\omega + ikv_x} \frac{df_0}{dv^2} d\mathbf{v} = 1 \qquad (2.60)$$

Let us look at two limits which illustrate the instability and the Jeans length. For small k we expand in kv_x/ω to get

$$1 \approx -\frac{8\pi G m}{k^2} \int \left[1 \frac{kv_x}{\omega} + \left(\frac{kv_x}{\omega}\right)^2\right] \frac{dF_b}{dv^2} d\mathbf{v} \qquad (2.61)$$

and the first term gives zero by symmetry and the second term (after integrating by parts and using $\rho_b = m \int F_b d\mathbf{v}$) gives

$$\omega = \sqrt{4\pi G \rho_b} \qquad (2.62)$$

This is the same gravitational instability shown earlier, in agreement with equation (2.47) for small k. If we set $\omega = 0$ in equation (2.60), we find the critical Jeans length $\lambda_J = 2\pi/k_J$. Equation (2.60) gives (after integration by parts)

$$\lambda_J = \langle v^2 \rangle^{-1/2} \sqrt{\frac{\pi}{G\rho_b}} \tag{2.63}$$

This is similar to the collisional expression (2.48), but now velocity dispersion behaves like the pressure. The stability at $\lambda < \lambda_J$ is not due to wave propagation, but is due to free streaming, related to Landau damping in Plasma physics.

As Zel'dovich (1965) points out, the nature of the instability in our expanding fluid model is well illustrated by the physics presented above, though we must include the important effects of universal expansion when we present the equations of motion and draw firm conclusions, as we do below. We may gain some insight by crudely modifying the ideas above to accommodate the fact that the background density depends on time. For example, let us pick Einstein–de-Sitter expansion; then the Friedman equation (2.4) gives $a(t) \sim t^{2/3}$ and in particular $\rho_b = (6\pi G t^2)^{-1}$. For $k \ll k_J$ (a very weak constraint in the matter dominated era), our plane wave perturbation solution (2.49) applies, with $e^{\omega t}$ time dependence, and $\omega = \pm\sqrt{4\pi G\rho_b}$ independent of k to good approximation, so a density perturbation could be represented as

$$\delta(\mathbf{x}, t) = e^{\omega t} \int \frac{d\mathbf{k}}{(2\pi)^3}\, \delta_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.64}$$

$$= A(t)\, \psi(\mathbf{x}) \tag{2.65}$$

with the space and time factors separated. To account for the time dependence we simply "integrate" $\dot{\delta} = \pm\omega\delta$, with $\omega = \sqrt{2/3}\,t^{-1}$, and we find that the time dependence[10] of a perturbation is $A(t) \sim t^{\pm\sqrt{2/3}}$. So perturbations grow by a power law, rather than by exponential growth, in an expanding universe. This turns out to be crucial, as there is a sort of agreement in timing between the instability and the cosmic expansion—the instabilities do not develop so quickly as to render the background model useless. This may even be somewhat obvious in hindsight, since there is no preferred time scale (the instability rate ω coincides with the expansion rate $\dot{a}/a$), so that by dimensional analysis $a(t)$ and $A(t)$ are scale free power laws. This not strictly true if, for example, $\Omega_0 < 1$, because there is a curvature scale, and eventually the instability $A(t)$ is quelled (basically because the universe dilutes too rapidly) when $z \lesssim \Omega_0^{-1}$. Just as important is the fact the wavelengths of the perturbations scale like $a(t)$, so that in the strict analysis, the $A(t)\psi(\mathbf{x})$ split holds true, but in comoving coordinates $\mathbf{x}$, so the growth of long wave perturbations is independent of $\mathbf{k}$ for small comoving k, and the perturbation retains its initial shape.

---

[10] The strictly correct conclusion, shown below, is that the growing and decaying solutions are $A \sim t^{2/3}$ and $A \sim t^{-1}$.

It is customary to introduce the Jeans mass M_J, which gives a lower bound on the objects which form. For a comoving λ_J this is defined as

$$M_J = \left(\frac{a\lambda_J}{2}\right)^3 \bar{\rho}_m \tag{2.66}$$

where $\bar{\rho}_m$ is the background matter density. Since the critical wavelength scales with $a(t)$ and $\bar{\rho}_m$ varies like $a(t)^{-3}$, M_J is independent of these variations and is a particularly convenient quantity.

To illustrate the cosmological Jeans mass, we note that the analogs to equations (2.29) and (2.30) in the full analysis give a similar equation for the evolution of a mode $\delta_{\mathbf{k}}$ ($\mathbf{k}$ is the comoving wavevector)

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} + \left(\frac{v_s^2 k^2}{a^2} - 4\pi G\bar{\rho}\right)\delta_{\mathbf{k}} = 0 \tag{2.67}$$

from which we draw similar conclusions concerning the comoving Jeans length λ_J

$$a(t)\lambda_J = v_s\sqrt{\frac{\pi}{G\bar{\rho}}} \tag{2.68}$$

The Jeans mass in a cosmological setting depends on the background energy density (which varies) and the nature of the matter (which also varies). Determining M_J in a detailed fashion is a somewhat tractable problem in cosmology (Peebles 1980 and references therein) but even so it is unclear how the instability further develops and how the various preferred length scales which do occur are ultimately related to galaxy or cluster formation.

To indicate rough orders of magnitude we show in Figure 2-2 the evolution of the baryon Jeans mass (in units of $M_\odot$) with time (in terms of $a/a_0 = (1+z)^{-1}$) covering the radiation and matter dominated eras. Details are presented in Weinberg (1972), Peebles (1980), and Kolb and Turner (1990). We now address the qualitative trends of M_J shown in Figure 2-2.

Before matter-radiation equality the Jeans mass increased as a^3, which we deduce from equation (2.66), since $\rho_m \sim a^{-3}$, $v_s^2 = c/3$, and the background radiation density is proportional to a^{-4}. Simply using the photon and neutrino background density from equation (2.21) and a baryon density of (somewhat arbitrarily) $\Omega_B h^2 \approx 0.05$ leads to the expression plotted in Figure 2-2 for $z \gtrsim 10^4$. Note that that this radiation dominated regime, v_s is essentially c and that the baryon Jeans mass is roughly the same as the total baryon mass within the horizon. That means that the perturbations that can grow are on superhorizon scales and behave as $\delta \sim t$ (Lifshitz 1946) according to the full relativistic treatment (cf. Peebles 1980, §86). For $10^4 \gtrsim z \gtrsim 10^3$, the sound speed dropped (but not abruptly). Since $v_s^2 \sim T \sim a^{-1}$ and we take $\rho \sim a^{-3}$, the Jeans mass is roughly independent of z here. After decoupling, the pressure support from matter, not photons, is small and the Jeans mass drops precipitously. We show the further decrease of M_J after decoupling varying as $a^{-3/2}$, just assuming that the nonrelativistic matter cools adiabatically (which implies $T \sim a^{-2}$), ignoring

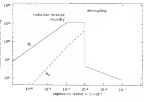


Figure 2·2 Schematic diagram showing the baryon Jeans mass M_J as a function of the expansion factor $a(t)/a(t_0)$ Also shown is the Silk mass M_S corresponding to the photon diffusion damping scale which is in effect before decoupling any reionization, etc., somewhat inaccurately since we know that the intergalactic medium is not such a quiescent system.

Prior to decoupling, there is a damping of baryonic fluctuations due to photon diffusion. The sub-Jeans fluctuations propagate like sound waves with the photons providing the pressure support. The diffusion of photons out of baryonic fluctuations effectively damps fluctuations below the diffusion length scale (Silk 1968). The mass scale M_S (known as the Silk mass) so defined provides a lower bound on structures as fluctuations with $M < M_S$ are essentially washed out. The effect increases toward decoupling as the diffusion increases, but the effect ceases after decoupling. Without the damping, photons are trapped in a perturbation by Thompson scattering (the γ e^- scattering makes for coupling with the protons too since protons and electrons are coupled electromagnetically even in the fully ionized plasma).

We can crudely estimate the magnitude and trend of the damping scale by estimating the diffusion length scale d and then taking $M_S \sim \rho_m d^3$. The photon mean free path is $\lambda_\gamma \sim (n_e \sigma)^{-1}$, where σ is the Thomson cross section and n_e is the electron number density (which is just that of the baryons, $n_e \sim \Omega_B h^2 a^{-3}$). The diffusion constant is $\sim \lambda_\gamma c$ and so the length scale is $d \sim (\lambda_\gamma c t)^{1/2}$. This gives $M_S \sim a^{9/2}$ during the radiation era and $M_S \sim a^{15/4}$ in the matter era prior to decoupling. Being a little more careful with the numerical coefficients, using an ionization fraction 0.1 at decoupling (Kolb & Turner 1990), we have

$$M_S = 6 \times 10^{12} (\Omega_0/\Omega_B)^{3/2} (\Omega_0 h^2)^{-5/4} M_\odot \tag{2.69}$$

at decoupling. Note that $M_S < M_J$ in Figure 2.2.

A universe with Ω_0 contributed by baryons only is practically ruled out, due primarily to problems relating to cosmic background radiation $\Delta T/T$ measurements and to primordial nucleosynthesis constraints. Given the structure we see today, the $\delta(\mathbf{x})$ values at decoupling would have to have been fairly large, corresponding background temperature anisotropies on the order of $\Delta T/T \approx 10^{-3} - 10^{-4}$. This has been a clear problem for years, given the decreasing upper bound constraints placed on $\Delta T/T$ by experimentalists, which finally led to a statistical detection of $\Delta T/T \approx 10^{-5}$ fluctuations (Smoot et al. 1992). The nucleosynthesis problem stems from the robust predictions of the light element abundances, which proved to be one of the experimental foundations of big bang models. The small baryon number of the universe is parametrized by the baryon to photon ratio $\eta = n_B/n_\gamma$. We can estimate the current η as follows. Since $n_B = \rho_c(t)\Omega_B/m_p$ and $n_\gamma = (\zeta(3)/\pi^2) g T^3$ with $\zeta(3) \approx 1.2$ and $T = 2.73\,\mathrm{K}$, we have (Kolb & Turner 1990)

$$\eta \approx 2.7 \times 10^{-8} \Omega_B h^2 \qquad (2.70)$$

The successes and constraints, e.g. $3 \times 10^{-10} \lesssim \eta \lesssim 10 \times 10^{-10}$, provided by the nucleosynthesis calculations seem to indicate that the dark matter we detect in halos and large structures must be (at least in part) nonbaryonic.

In the case of nonbaryonic dark matter, structure formation is qualitatively similar to the baryon scenario briefly described above, but with significant differences in that the dark matter models are able to overcome the problems of the baryon only models. However, there are many different dark matter scenarios, and predictions tend to be very model-dependent, and no particular model can claim champion status.

We note some general features of a theory in which nonbaryonic species with nonzero mass decouple early yet remain cosmologically significant. Details, including some of the particle physics background, are covered by Kolb and Turner (1990). The collisionless Jeans mass, from the Jeans length of equation (2.63), still pertains, with the velocity dispersion $\langle v^2 \rangle$ being the analog to the sound speed due to pressure (though pressure is not the right term in the collisionless case of WIMPS—weakly interaction massive particles). Damping due to free streaming is effective on scales smaller than the Jeans length (with $M_{FS} < M_J$ in general, in analogy to the case of photon diffusion damping) so strictly speaking one would speak of M_{FS} rather than M_J A full analysis (Bond & Szalay 1983) shows that there are two chief scales which occur λ_{FS} and d_{hor} at the epoch of matter radiation equality though these length scales are not generally sharply defined in the "initial" spectrum of density fluctuations $\delta(\mathbf{k})$ at high redshift ($z \approx 1000$) Decoupling does not influence matters at all in contrast to the baryon case Gravitational instability is free to operate in the nonbaryonic component before decoupling which overcomes the timing or $\Delta T/T$ problems of the baryon only model If the dark matter fluctuations have grown by the time decoupling occurs then the baryons fluctuations (already damped on small scales) begin growing, and in time they mimic the nonbaryonic fluctuations if the WIMPS provide the dominant source of gravitation

These models vary widely and they generally are susceptible to some sort of experimental verification For example, the simplest Hot Dark Matter (HDM) and Cold Dark Matter (CDM) seem to be ruled out If we ignore the hydrodynamics of collapse and galaxy formation (which cannot be ignored in general) then the various models properties show up in the spectrum of small initial density fluctuations, when $\gtrsim 1000$

2 2 3 Equations of Motion

We study perturbations to the homogeneous expanding spacetime with line element

$$ds^2 = dt^2 - a^2(t)\, d\mathbf{x}^2 \tag{2 71}$$

The full machinery of general relativity is not needed in our study as we confine ourselves to the slow motion subhorizon limits The Newtonian weak field limit has a line element which we may write (e g for a spatially flat background) as

$$ds^2 = (1 + 2\psi)\, dt^2 - (1 - 2\phi)\, a^2(t)\, d\mathbf{x}^2 \tag{2 72}$$

where φ is the Newtonian gravitational potential coupled to the density fluctuations via the Poisson equation, as we describe below. The $\varphi \ll c^2$ constraint is essential to the validity of the limit, but this does not at all preclude large values of $\delta(\mathbf{x}, t)$, the density contrast field. The line element (2.71) about which we perturb can be open, closed, or flat ($\Omega = 1$), though we certainly use approximations based on $\Omega \approx 1$, for example when we use the $e^{i\mathbf{k}\cdot\mathbf{x}}$ mode expansion.

The equations of motion in the fluid model will be the direct analogs of the usual continuity, Euler, Poisson, and pressure $p(\rho)$ equations, modified to apply in an expanding universe in comoving coordinates $\mathbf{x}$. Before we write out those equations, let us separate out from the equations for a mass m in an expanding background the parts pertaining to the expansion, thus leaving the perturbed motion.

The proper coordinates $\mathbf{r}$ and comoving coordinates $\mathbf{x}$ are related by

$$\mathbf{r} = a(t)\mathbf{x} \tag{2.73}$$

$$\dot{\mathbf{r}} = \dot{a}\mathbf{x} + \mathbf{v}(\mathbf{x}, t) \tag{2.74}$$

where the peculiar velocity $\mathbf{v} = a\dot{\mathbf{x}}$ has been separated from the Hubble flow. The time-time component of the Einstein equations and the geodesic equation, in the Newtonian limit, subsuming the expansion and using $\mathbf{r}$ coordinates, are the familiar

$$\nabla_r^2 \Phi = 4\pi G \rho \tag{2.75}$$

$$\ddot{\mathbf{r}} = -\nabla_r \Phi \tag{2.76}$$

where the density is $\bar{\rho}(1 + \delta)$. We neglect pressure and cosmological constant, which could in principle be included; such effects are included in Peebles (1980). Separating out the part of Φ due the expansion (using equations (2.75) and (2.5), which follows here from (2.76) also), we write

$$\Phi = \varphi(\mathbf{x}, t) + \frac{2}{3}\pi G \bar{\rho} a^2 = \varphi(\mathbf{x}, t) - \frac{1}{2} a\ddot{a} x^2 \tag{2.77}$$

Thus we have the Poisson equation written in comoving coordinates

$$\nabla^2\varphi(\mathbf{x},t) = 4\pi G\rho a^2\delta(\mathbf{x},t) \tag{2.78}$$

which couples the density fluctuations to the perturbations in the expanding background metric. For brevity, we denote $\nabla_{\mathbf{x}}$ by ∇.

All this is more succinctly noted (cf. Peebles 1980) if we start with the action S

$$\begin{aligned} S &= -m\int d\tau = -\int\sqrt{g_{\mu\nu}dx^{\mu}dx^{\nu}} \\ &\approx \int dt\left[\frac{1}{2}m(a\dot{\mathbf{x}}+\dot{a}\mathbf{x})^2 - m\Phi\right] + \text{const} \end{aligned} \tag{2.79}$$

Then we let the usual $L = T - V$ as we have above go to $L - \psi$ with a canonical transformation $\psi = \frac{1}{2}ma\dot{a}x^2$. This transformation makes good the separation between the Hubble flow and peculiar motions, giving

$$L = \frac{1}{2}ma^2\dot{\mathbf{x}}^2 - m\varphi \tag{2.80}$$

As noted above, φ and δ are coupled by the Poisson equation (2.78). The Euler Lagrange equations from (2.80) give

$$\frac{d\mathbf{p}}{dt} = \frac{d}{dt}(ma^2\dot{\mathbf{x}}) = -m\nabla\varphi \tag{2.81}$$

which we can rewrite as

$$\frac{d\mathbf{v}}{dt} + \frac{\dot{a}}{a}\mathbf{v} = -\frac{1}{a}\nabla\varphi \tag{2.82}$$

We now write the equations of motion of self gravitating matter in an expanding universe. Since we are able to separate out the background Hubble flow, utilizing $\rho(t) \sim a^{-3}$ and the Friedman equation in the physical $\mathbf{r}$ coordinates, the equations in our fluid model will be written in terms of $\delta(\mathbf{x},t)$ and $\mathbf{v}(\mathbf{x},t)$ in comoving coordinates $\mathbf{x}$.

There are two approaches (Peebles 1980, §9) which lead to the same fluid equations: the ideal fluid and collisionless matter approaches.

If we consider the particle mean free path to be small and that the universe is filled with a self gravitating Newtonian ideal fluid then we may begin with the usual continuity and Euler equations in r coordinates. Then if we change coordinates from r to $\mathbf{x} = \mathbf{r}/a$ and follow through with the complete change of variable procedure for the density $\rho(\mathbf{r}/a, t)$ and velocity $\dot{\mathbf{r}} = H\mathbf{r} + \mathbf{v}(\mathbf{r}/a, t)$ the fluid equations become

$$\frac{\partial \rho}{\partial t} + 3\frac{\dot{a}}{a}\rho + \frac{1}{a}\nabla\cdot\rho\mathbf{v} = 0 \qquad (2.83)$$

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{a}\mathbf{v}\cdot\nabla\mathbf{v} + \frac{\dot{a}}{a}\mathbf{v} = -\frac{1}{\rho a}\nabla p - \frac{1}{a}\nabla\varphi \qquad (2.84)$$

Note that we can rewrite (2.83) using $\delta(\mathbf{x}, t)$ as $a\dot{\delta} = -\nabla\cdot[(1+\delta)\mathbf{v}]$.

The collisionless approach is more natural for the development of large structures. We have already separated out the expansion in the single particle dynamical equations above. The smooth gravitational potential $\varphi(\mathbf{x}, t)$ determines $\delta(\mathbf{x}, t)$ describing noninteracting (but for gravity) fluid elements of mass m (which could be dark matter or galaxies, for example). The one particle distribution function is $f(\mathbf{x}, \mathbf{p}, t)$ where the canonical momentum is $\mathbf{p} = ma^2\mathbf{v}$ as in equation (2.81). The background density is $m\bar{n}$ so the comoving number density $\bar{n}a^3$ is constant. The first three moments of $f(\mathbf{x}, \mathbf{p}, t)$ are defined as

$$\int d^3p\, f = \bar{n}a^3(1+\delta) \qquad (2.85)$$

$$\int d^3p\, p^i f = \bar{n}a^3(1+\delta)mav^i \qquad (2.86)$$

$$\int d^3p\, p^i p^j f = \bar{n}a^3(1+\delta)m^2a^2v^iv^j \qquad (2.87)$$

The dynamics of f is specified by the Vlasov equation

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{ma^2}\nabla f - m\nabla\varphi\cdot\nabla_{\mathbf{p}} = 0 \qquad (2.88)$$

Taking moments of the Vlasov equation gives the continuity and Euler equations for collisionless fluid in an expanding background

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot[(1+\delta)\mathbf{v}] = 0 \tag{2.89}$$

$$\frac{\partial\mathbf{v}}{\partial t} + \frac{1}{a}\mathbf{v}\cdot\nabla\mathbf{v} + \frac{\dot{a}}{a}\mathbf{v} = -\frac{1}{a}\nabla\varphi \tag{2.90}$$

These equations, combined with

$$\nabla^2\varphi(\mathbf{x},t) = 4\pi G\bar{\rho}a^2\delta(\mathbf{x},t), \tag{2.91}$$

are the desired equations of motion. We can combine (2.89) and (2.90) to obtain an important equation governing density fluctuations (Peebles 1980)

$$\frac{\partial^2\delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial\delta}{\partial t} - 4\pi G\bar{\rho}\delta = 4\pi G\bar{\rho}\delta^2 + \frac{1}{a^2}\nabla\varphi\cdot\nabla\delta + \frac{1}{a^2}\nabla_i\nabla_j[(1+\delta)v^iv^j] \tag{2.92}$$

Our approach will be to use perturbation theory based on these equations.

2.2.4 Linear Perturbation Theory

We begin here by laying out the simplest, yet important, case of linear perturbation theory. Early works by Doroshkevich and Zel'dovich (1963) and Zel'dovich (1965) clarified the nature of perturbation theory in the Newtonian approximation. Papers which went beyond first order theory were supplied by Zel'dovich (1970), Peebles (1980), Fry (1984), and Goroff, Grinstein, Rey and Wise (1986).

If we keep only the lowest order terms in δ and $\mathbf{v}$ in the fluid equations, we have as our system the linearized continuity and Euler equations

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot\mathbf{v} = 0 \tag{2.93}$$

$$\frac{\partial\mathbf{v}}{\partial t} + \frac{\dot{a}}{a}\mathbf{v} = -\frac{1}{a}\nabla\varphi \tag{2.94}$$

along with the Poisson equation For the evolution of δ we need only keep the left hand side of (2 92) which gives us

$$\frac{\partial^2\delta}{\partial t^2} + 2\frac{a}{a}\frac{\partial\delta}{\partial t} - 4\pi G\rho\delta = 0 \tag{2 95}$$

The unperturbed configuration $\mathbf{v} = 0$ $\delta = 0$ $\varphi = 0$ is now an actual solution to the equations of motion in contrast to the case of a static background for which we performed the Jeans analysis in §2 2 2 Note that if we include pressure as we wrote in the ideal fluid equation (2.84) and expand in $e^{i\mathbf{k}\cdot\mathbf{x}}$ modes with physical wavelength $2\pi a/k$ we arrive at the expanding Jeans analysis equation (2 67) advertised previously

Now we write the leading order terms in the perturbation theory expansion First note that the δ evolution equation (2 95) is completely local, and the solution separates immediately to

$$\delta(\mathbf{x}\ t) = A(t)\,\delta_0(\mathbf{x}) \tag{2 96}$$

So any initial fluctuations, $\delta_0(\mathbf{x}) = \delta(\mathbf{x}, t_i)$ at some early time t_i, in comoving coordinates $\mathbf{x}$ retain their shape and grow in amplitude The $A(t)$ solutions to

$$A + 2\frac{a}{a}A - 4\pi G\rho A = 0 \tag{2 97}$$

are covered thoroughly by Peebles (1980) For the $\Omega_0 = 1$ Einstein–de-Sitter case, or early when it is effectively valid $1 + z \gg |\Omega_0^{-1} - 1|$ and the curvature term in the Friedman equation (2.4) is negligible, we have $a \sim t^{2/3}$ and $6\pi G\bar{\rho}t^2 = 1$ so that equation (2 97) reduces to

$$A + \frac{4}{3t}A - \frac{2}{3t^2}A = 0 \tag{2 98}$$

which has two power law solutions

$$A(t) \sim t^{2/3} \quad A(t) \sim t^{-1} \tag{2 99}$$

So $\delta(\mathbf{x}, t)$ is the sum of two independent solutions, one growing and the other decaying, with the particular linear combination depending on initial conditions. We only focus on the $t^{2/3}$ solution, which soon dominates the t^{-1} term and leads to structure formation.

As we did in the Jeans analysis, we can split the peculiar velocity into curl free (longitudinal) and divergence-free (transverse) pieces. Only the longitudinal part is coupled to gravity in the Euler equation. Since a potential force term cannot generate vorticity, we have, in analogy to Kelvin's conservation of circulation in static a background, that the transverse piece $\mathbf{v}_\perp$ and the vorticity $\nabla \times \mathbf{v}_\perp$ decay. This can be seen from the curl of the linearized Euler equation, as only the transverse piece survives

$$\frac{\partial}{\partial t}(a\nabla \times \mathbf{v}_\perp) = 0 \qquad (2.100)$$

which implies $\mathbf{v}_\perp = \nabla \times \mathbf{F}(\mathbf{x})/a$ for some $\mathbf{F}$. We arrived at the conclusion from linear theory, but it is true in general, as (2.100) is reproducing the Kelvin result that

$$\Gamma = a \oint \mathbf{v} \cdot d\mathbf{x} \qquad (2.101)$$

is constant around a material loop. No vorticity can be generated in the large. Some dissipative small scale processes will do so. Similar conclusions hold in the purely collisionless case, but the vorticity on small scales occurs when the trajectories begin mixing (Peebles 1980, Zel'dovich and Novikov 1983). Consequently, we concern ourselves with potential flow for $\mathbf{v}$.

The linear $\mathbf{v}$ solution follows simply once we know that $\delta = A(t)\delta_0(\mathbf{x})$ and we can write the velocity as a gradient

$$\mathbf{v} = -\frac{1}{a}\nabla\varphi_v \qquad (2.102)$$

From the continuity equation $a\dot{\delta} = -\nabla \cdot \mathbf{v}$ and $\dot{\delta}/\delta = \dot{A}/A$, we have $\nabla^2\varphi_v \sim \delta$ so $\varphi_v \sim \varphi$ up to time dependent factors. Getting the factors correct from the Poisson equation (2.81), we have

$$\mathbf{v} = -\frac{\dot{A}}{A}\frac{1}{4\pi G\rho a}\nabla\varphi \tag{2.103}$$

So we see that in linear theory $\mathbf{v}$ is parallel to the peculiar acceleration

$$\mathbf{g} = -\frac{1}{a}\nabla\varphi \tag{2.104}$$

As for time dependence, with the first order $\delta = A(t)\delta_0$ solution as the source in the Poisson equation, we see that the linear gravitational potential φ has time dependence

$$\varphi \sim \frac{A(t)}{a(t)}, \tag{2.105}$$

so it is nearly constant in time. For $\Omega = 1$ the linear potential is in fact time independent. The peculiar velocity solution has $\mathbf{v} \sim a\dot{A}$ and for $\Omega = 1$ (2.103) is simply $\mathbf{v} = \mathbf{g}\,t$ (so $\mathbf{v} \sim t^{1/3}$ in that case).

The expressions involving A may be rewritten (Peebles 1980) by defining

$$f(\Omega) = \frac{d\log A}{d\log a} = \frac{\dot{A}a}{A\dot{a}} \tag{2.106}$$

which characterizes the $A(t)$ solutions conveniently. For $\Omega = 1$ we have $f = 1$. For Ω asymptotically near 1, $f \approx \Omega^{4/7}$ and $f \approx \Omega^{0.6}$ is a good numerical approximation for smaller Ω. The expressions for the linear $\mathbf{v}$ and δ can thus be written as

$$\mathbf{v} = -\frac{2f}{3Ha\Omega}\nabla\varphi \tag{2.107}$$

and

$$\delta = -\frac{A}{\dot{A}a}\nabla \cdot \mathbf{v} = -\frac{1}{faH}\nabla \cdot \mathbf{v} \tag{2.108}$$

The general φ solution to the Poisson equation may be written as

$$\Delta = -\frac{1}{4\pi}\int d^3x'\,\frac{\delta(\mathbf{x}',t)}{|\mathbf{x}-\mathbf{x}'|} \tag{2.109}$$

where Δ is a rescaled gravitational potential

$$\nabla^2\Delta = \delta \tag{2.110}$$

$$\varphi = 4\pi G\bar{\rho}a^2\Delta \tag{2.111}$$

The Poisson solution is used self consistently order by order in perturbation theory. In linear theory, the rescaled gravitational potential with time dependence $\Delta \sim A(t)$ is related to δ by (2.109) with the first order $\delta = A(t)\delta_0$ solution as the source.

The linear theory presented here is strictly valid only for $\langle\delta^2\rangle \ll 1$ and on large (yet subhorizon) scales the simplifications made here—using Newtonian limit, neglecting non-gravitational physics—are appropriate. It would seem that current large-scale fluctuations would retain some of the information from the linearly evolved quantities. At the least, we have, for example, that an initial ($t = t_i$) perturbation spectrum will evolve in time like $A^2|\delta(\mathbf{k}, t_i)|^2$ for small k (large scales). In fact, going beyond simple first order theory seems to be more fruitful than might at first be expected (Zel'dovich 1970, Fry 1984). Large-scale statistics in the galaxy distribution seem to be related to the dynamics provided by perturbation theory. The relationship is difficult to untangle in practice, as real observational data rarely have a clean interpretation.

2.3 Statistics

2.3.1 Correlations

In an attempt to characterize the large-scale matter distribution, we utilize cosmological statistics which are in principle measurable. The matter density ρ is directly related, we assume, to the galaxy number density n_g, and we use the dimensionless density contrast fields

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \tag{2.112}$$

$$\delta_g(\mathbf{x}) = \frac{n_g(\mathbf{x}) - \bar{n}_g}{\bar{n}_g} \tag{2.113}$$

The peculiar velocity field $\mathbf{v}(\mathbf{x})$ is measurable in principle if it is a gradient. Since $\mathbf{v}$ is fundamentally related to $\delta(\mathbf{x})$ and not the potentially biased δ_g, statistics involving velocities may have great value in the future, but current data are only preliminary. We focus on the density. From the equations of motion, the φ, $\mathbf{v}$, and δ fields are all amenable to study.

Confining our observational interest to regions $\lesssim 1000\,h^{-1}$ Mpc, we are effectively ignoring lookback times. We somewhat cavalierly treat galaxies as simple objects which trace the mass, with the possibility of a simple local biasing relation $\delta_g = f(\delta)$. The statistics of galaxy properties and their evolution is an entire, related field of study and is reviewed by Kron (1995). The observable density field naturally uses physical coordinates, so we could write the measured density as $\delta(\mathbf{x}_{phys})$. We may take $\mathbf{x}_{phys}$ as coinciding with the comoving coordinates at the present time if $\mathbf{x}_{phys} = a\mathbf{x}$ and $a(t_0) = 1$ presently.

One of the central problems surrounding the data stems from the immense difficulties in amassing deep galaxy surveys. The photometric selection criteria and the luminosity distributions for the survey galaxies can in principle be accounted for by a selection function which gives various statistical weights to the galaxies included in statistical estimators involving δ_g. We do not include such weights in formulas below (they are included implicitly when any statistic is to be estimated from real data).

There is a fairly rich history of attempts to use statistical measures on the galaxy data, and this is covered by Peebles (1980). As the quality and quantity of data improved in the 1960's and 1970's, correlation statistical methods were pioneered by

Peebles and his collaborators. Counts-in-cell statistics, the void probability function in particular, have proved useful. Statistics related to shape, topology and percolation should prove useful, though no standard has really taken hold.

Note that the initial conditions on $\delta(\mathbf{x}, t_i)$ are random in nature, usually given by the spectrum of perturbations in Fourier space. Early universe theories such as inflation can specify the statistics of the primordial fluctuations (and this is one of their main selling points as theories). Predictions concerning structure formation are statistical. We do not wish to predict the location of the Coma cluster in the Great Wall, or predict the exact origin of the local supercluster from initial conditions. Some conclusions are intuitively clear. For example, if a model is not able to produce "enough" large voids and filaments in its modeled galaxy distributions, then the theory is lacking. The statistics we study are similar to correlated noise, such as what we might study in investigating the choppy surface of water in a swimming pool or the random peaks and troughs in the ocean's water surface, assuming the coherence length of the noise is sampled many times over in our survey of the random water surface.

We consider $\delta(\mathbf{x})$ to be a homogeneous and isotropic random field. This is based on the Cosmological Principle, along with an ergodic hypothesis, whereby a large enough sample in our local universe (in the NGP direction, say) is an equivalent realization of the same random process as compared to a survey in the SGP direction. This is formalized by Peebles (1980) and is called the Fair Sample Hypothesis. The Slice of the Universe survey, Figure 1-9, is not quite a fair sample. The new LCRS slices, Figure 1-10, seem to show the requisite large-scale homogeneity.

If we take a large volume V which we assume to be a fair sample, then we define the important 2-point correlation function as

$$\xi(r) = \langle \delta(\mathbf{x}+\mathbf{r})\delta(\mathbf{x}) \rangle \,, \tag{2.114}$$

where the $\langle\ \rangle$ indicates an average over V (which we take as equivalent to an ensemble average). By homogeneity and isotropy, the dependence is on $r = |\mathbf{r}|$.

As we have introduced it, $\xi(r)$ is the autocorrelation function of a continuous density field $\delta(\mathbf{x})$. If we take the continuous field as our starting point (as the fundamental realization of the random process), then we may invoke a Poisson model to get a discrete realization of $\delta(\mathbf{x})$, which gives

$$dP = n^2[1 + \xi(r)]dV_1 dV_2 \tag{2.115}$$

as the joint probability that two objects are found in volume elements dV_1, dV_2 separated by r. Or we may start with a fundamentally discrete model which defines $\xi(r)$ by (2.115). The continuous and discrete approaches are related throughout the correlation hierarchy, as described by Peebles (1980), Bertschinger (1992), and Fry (1997).

Equation (2.115) has a nice conditional probability interpretation: if we choose an object at random from the realization, the probability that a neighbor resides at distance r from it at volume dV is

$$dP = \bar{n}dV[1 + \xi(r)]. \tag{2.116}$$

This gives a way to directly estimate ξ from galaxy counts, by comparing the actual counts to those of a random catalog. A robust result for $r \lesssim 10\,h^{-1}$ Mpc is that

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \tag{2.117}$$

with $r \approx 5\,h^{-1}$ Mpc and $\gamma \approx 1.8$, while for larger scales ξ is not well measured. The APM angular 2-point function of $\approx 2 \times 10^6$ galaxies gave a deprojected $\xi(r)$ in agreement with (2.117), as did the earlier Lick Observatory "million galaxies" counts. Though the APM ξ decreased at scales $\gtrsim 10\,h^{-1}$ Mpc, the amplitude was large in comparison with basic CDM predictions, for example.

The N-point correlation function is given by

$$dP = \bar{n}^N dV_1 \ldots dV_N [1 + N\text{-point function}] \qquad (2.118)$$

with a probabilistic interpretation as in equation (2.115). In particular, the 3-point function is given by

$$dP = \bar{n}^3 dV_1 dV_2 dV_3 [1 + \xi_{12} + \xi_{23} + \xi_{31} + \zeta_{123}], \qquad (2.119)$$

where $\xi_{12} = \xi(|\mathbf{x}_1 - \mathbf{x}_2|)$, and ζ_{123} depends on three numbers, the sides the $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ triangle defined by the three points. The connected, or cumulant, 3-point function is ζ, and it is this which is usually simply called the 3-point correlation function in cosmology. A similar expression (Fry & Peebles 1978) defines the reduced 4-point function in cosmology. In general, as in (2.119), the N-point function is a combinatoric sum over lower correlations ending with a final, connected N-point term. The correlation hierarchy and the underlying mathematical details are thoroughly covered by Fry (1997), Bertschinger (1992), Peebles (1980), and White (1978).

The third moment of δ gives

$$\langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \delta(\mathbf{x} + \mathbf{s}) \rangle = \zeta(r, s, |\mathbf{r} - \mathbf{s}|), \qquad (2.120)$$

and we see that ζ is the random field analog to the skewness of of a probability distribution function. The 4-point reduced correlation function η comes from the fourth moment of δ

$$\eta_{1234} = \langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle - \langle \delta_1 \delta_2 \rangle \langle \delta_3 \delta_4 \rangle - \langle \delta_1 \delta_3 \rangle \langle \delta_2 \delta_4 \rangle - \langle \delta_1 \delta_4 \rangle \langle \delta_2 \delta_3 \rangle \qquad (2.121)$$

so η is the analog of the excess kurtosis.

Direct estimations of the N-point galaxy correlations, Peebles and Groth (1975) for the 3-point and Fry and Peebles (1978) for the 4-point, show a hierarchical form

ry 1984) whereby higher order correlations are constructed from the 2 point func-
on. The simplest case, which is observationally robust, is that

$$\zeta_{123} \approx Q\,(\xi_{12}\xi_{23} + \xi_{23}\xi_{31} + \xi_{31}\xi_{12}) \tag{2.122}$$

for $r \lesssim 3\,h^{-1}$ Mpc, with $Q \approx 1$.

We now consider counts-in-cells statistics, which we relate to integrals over the correlation functions. Let us randomly place a cell of fixed size ($\approx 8\,h^{-1}$ Mpc in dimension for example) and shape repeatedly throughout a galaxy distribution. The counts N in the cell of volume V will be our random variable. The mean counts per cell is $\bar{N} = n_g V$, where n_g is the number density of galaxies. By analyzing the cell counts in V, further chopping V into infinitesimal volumes and using the ξ and ζ point process definitions (2.115) and (2.119), we obtain the following expressions for the moments of the N frequency distribution (Peebles 1980)

$$\Delta \equiv \frac{N - \bar{N}}{\bar{N}} \tag{2.123}$$

$$\langle \Delta^2 \rangle = \frac{1}{V^2}\int d^2V\xi + \frac{1}{\bar{N}} \tag{2.124}$$

$$\langle \Delta^3 \rangle = \frac{1}{V^3}\int d^3V\zeta + 3\frac{\langle \Delta^2 \rangle}{\bar{N}} - \frac{2}{\bar{N}^2} \tag{2.125}$$

$$\langle \Delta^4 \rangle - 3\langle \Delta^2 \rangle^2 = \frac{1}{V^4}\int d^4V\eta + 6\frac{\langle \Delta^3 \rangle}{\bar{N}} - 11\frac{\langle \Delta^2 \rangle}{\bar{N}^2} + \frac{6}{\bar{N}^3} \tag{2.126}$$

Apart from the discreteness, or shot noise, terms, we see that the moments of counts give correlation functions integrated over a volume V.

2.3.2 Fourier Spectra

It is often most convenient to work in the Fourier domain. For the transform pair of a continuous density field, we have

$$\delta(\mathbf{k}) = \int d^3x\, e^{-i\mathbf{k}\cdot\mathbf{x}}\,\delta(\mathbf{x}) \tag{2.127}$$

$$\delta(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\, e^{i\mathbf{k}\cdot\mathbf{x}}\,\delta(\mathbf{k}) \tag{2.128}$$

The Dirac delta function is represented by

$$\delta_D(\mathbf{k}) = \frac{1}{(2\pi)^3} \int d^3x\, e^{\pm i\mathbf{k}\cdot\mathbf{x}} \tag{2.129}$$

Since $\delta(\mathbf{x})$ is real (naturally), we have the for the complex $\delta(\mathbf{k})$ field

$$\delta(-\mathbf{k}) = \delta(\mathbf{k})^* \tag{2.130}$$

If $\delta(\mathbf{x})$ is a realization of a homogeneous, isotropic random process with 2 point spatial correlations $\xi_{12} = \langle\delta_1\delta_2\rangle$, then we may compute the second moment of $\tilde{\delta}$ to be

$$\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\rangle = \left[(2\pi)^3\delta_D(\mathbf{k}_1 + \mathbf{k}_2)\right] P(k_1) \tag{2.131}$$

where $P(k)$ is the power spectrum, which forms a transform pair with $\xi(r)$,

$$P(k) = \int d^3x\, e^{-i\mathbf{k}\cdot\mathbf{x}} \xi(|\mathbf{x}|) \tag{2.132}$$

In this second moment calculation, we see that the "momentum conserving" delta function comes from translation invariance, as $\langle\delta_1\delta_2\rangle$ depends on $\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2$ only, and that isotropy (the $|\mathbf{r}|$ dependence of ξ) likewise gives the isotropic k dependence of $P(k)$. Thus we see that an average of $\delta(\mathbf{k})$ in k-space, at fixed k, gives the power spectrum of perturbations, or 2 point correlations, of the random process in the sense that

$$P(k) \sim \langle|\delta(\mathbf{k})|^2\rangle \tag{2.133}$$

We need to relate the continuous transform analysis to the practical discrete case which we use later on in calculations and simulations. First, note that there is no equality in (2.133), as according these definitions, $P(k)$ is a power spectral density with units of volume, while $\langle|\delta(\mathbf{k})|^2\rangle$ is the power per mode, with dimensions of volume squared. In our numerical simulations, and in analyzing simulated or real data, we go to a discrete case whereby data is contained in a large fundamental cube of volume $V = L^3$, on which we impose periodic boundary conditions. This simply invokes the usual Fourier analysis whereby a discrete Fourier transform corresponding to a Fourier series goes over to the continuous Fourier transform in the large V limit. We shall place an $n \times n \times n$ grid in x space with grid spacing $\Delta x = L/n$. This grid remains in place in k-space with grid spacing $\Delta k = 2\pi/L$ and there the relation is

$$\langle|\delta_d(\mathbf{k})|^2\rangle \approx P(k)\,(\Delta k)^3 \tag{2.134}$$

where the $\delta_d(\mathbf{k})$ in this equation is the discrete transform

$$\delta(\mathbf{k}) \approx \delta_d(\mathbf{k})(\Delta k)^{-3} \tag{2.135}$$

The discrete transform pair is with $\mathbf{x}$ and $\mathbf{k}$ lying on the n^3 grid points.

$$\delta_d(\mathbf{k}) = \frac{1}{n^3}\sum_{\mathbf{x}} e^{-i\mathbf{k}\cdot\mathbf{x}}\delta_d(\mathbf{x}) \tag{2.136}$$

$$\delta_d(\mathbf{x}) = \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}\delta_d(\mathbf{k}) \tag{2.137}$$

If we wish to make the $\int \leftrightarrow \sum$ correspondence seem more transparent we may introduce V (formally $V = \int d^3x$) into the measure, giving a dimensionless $\delta(\mathbf{k})$ as follows,

$$\delta(\mathbf{k}) = \int \frac{d^3x}{V} e^{-i\mathbf{k}\cdot\mathbf{x}}\,\delta(\mathbf{x}) \tag{2.138}$$

$$\delta(\mathbf{x}) = \int \frac{V\,d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}}\,\delta(\mathbf{k}) \tag{2.139}$$

Then the Dirac delta function (2.129) corresponds to the discrete Kronecker delta from

$$\sum_{\mathbf{x}} e^{\pm i\mathbf{k}\cdot\mathbf{x}} = n^3\delta_D(\mathbf{k}) \tag{2.140}$$

The second moment is then

$$\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\rangle = \left[\frac{(2\pi)^3}{V}\delta_D(\mathbf{k}_1 + \mathbf{k}_2)\right]P(k_1) \tag{2.141}$$

where the [ ] term is 0 or 1 and the notation

$$P(k) \equiv \langle|\delta(\mathbf{k})|^2\rangle \tag{2 142}$$

as used in this equation refers to a dimensionless power per mode which changes (as it should not if it were a true power spectrum) with the grid size The correspondence with a true power spectrum is understood and anyway nothing is really wrong since an overall normalization will fix the scale (if one changes grid size for example) This might be a little sloppy but it is useful There is really no problem as long as one is thinking in terms of one or the other—either continuous or discrete—during a calculation We go ahead with the notation in (2 142) As we see beginning in Chapter 4 it is thus $P(k)$ which is the immediate raw data (apart from possible discreteness and selection function effects) from a discrete Fourier analysis of data. The definitions of the power spectrum and related quantities are clarified by Bertschinger (1992)

From the the Fourier transform pair (2 138) and (2 139) we see that

$$\sigma^2 = \langle\delta(\mathbf{x})\delta(\mathbf{x})\rangle = \int \frac{V\, d^3k}{(2\pi)^3} P(k) \tag{2 143}$$

and so the contribution to the local variance σ^2 from the interval $d\ln k = dk/k \sim 1$ is $\Delta^2(k) = Vk^3P(k)/2\pi^2$ Furthermore the variance $\langle\delta^2\rangle_R$ at length scale $R \sim k^{-1}$ is $\sim k^3P(k)$ which shows part of the intuitive appeal of $P(k)$ as a description This can be seen from filtering considerations Define a window function W with an effective width R in x-space and let δ_W be the smoothed, or convolved, field.

$$\begin{aligned} \delta_W(\mathbf{x}) &= \int \frac{d^3x'}{V} W(x')\delta(\mathbf{x}-\mathbf{x}') \\ \int \frac{d^3x}{V} W(x) &= 1 \end{aligned} \tag{2 144}$$

For example we have the x space top hat

$$\begin{aligned} W(r) &= \frac{3V}{4\pi R^3}, \ (r < R) \\ W(k) &= \frac{3}{y^3}(\sin y - y\cos y) \quad y = kR \end{aligned} \tag{2 145}$$

and we could use a Gaussian filter to minimize the high frequency sharp edge effects if need be. The filtered moment works out as

$$\sigma_R^2 = \langle \delta_W(\mathbf{x})\delta_W(\mathbf{x})\rangle = \int \frac{V d^3k}{(2\pi)^3} P(k)|W(k)|^2 \tag{2.146}$$

and from this we have $\sigma_R \sim \Delta(k = R^{-1})$, though this depends on the form of $P(k)$, e.g. for $P \sim k^n$ we need $-3 < n < 1$ (cf. Kolb & Turner 1990).

The bispectrum and trispectrum (Peebles 1980, Baumgart and Fry 1991) can be defined from the third and fourth moments in k-space, analogously to equations (2.141) and (2.142). We now compute the third k-space moment $\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle$ clarifying the relation between the bispectrum and the spatial 3-point function

$$\begin{aligned}\langle\tilde{\delta}_1\tilde{\delta}_2\tilde{\delta}_3\rangle &= \int \frac{d^3x_1}{V}\frac{d^3x_2}{V}\frac{d^3x_3}{V}\langle\delta_1\delta_2\delta_3\rangle e^{-i\sum \mathbf{k}_i\cdot\mathbf{x}_i} \\ &= \int \frac{d^3r}{V}\frac{d^3s}{V}\zeta(r, s, |\mathbf{r}-\mathbf{s}|)e^{-i(\mathbf{k}_2\cdot\mathbf{r}+\mathbf{k}_3\cdot\mathbf{s})}[\delta_D] \\ &= [\delta_D]\, B(k_1, k_2, k_3)\end{aligned} \tag{2.147}$$

In the second step we changed coordinates $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \mapsto \mathbf{x}_1, \mathbf{x}_1 + \mathbf{r}, \mathbf{x}_1 + \mathbf{s})$ and utilized the homogeneity of ζ. The notation $[\delta_D]$ refers to the momentum conserving delta function

$$[\delta_D] \equiv \left[\frac{(2\pi)^3}{V}\delta_D\left(\sum \mathbf{k}_i\right)\right] \tag{2.148}$$

which in this case gives 1 for closed triangles on the $\mathbf{k}$ grid and gives 0 otherwise. Thus we have the convenient notation, similar to (2.142), which says

$$B(k_1, k_2, k_3) \equiv \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle\Big|_{\sum \mathbf{k}_i = 0} \tag{2.149}$$

The hierarchical form for ζ_{123}, equation (2.122), goes over in k space to (Fry & Seldner 1982)

$$B_{123} \approx Q\,(P_1P_2 + P_2P_3 + P_3P_1) \tag{2.150}$$

with the same Q value, if Q is constant. In (2.150), we have $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$ and $P_1 = P(k_1)$. In practice, the power P and bispectrum B are useful on both small

$r \lesssim 10\,h^{-1}$ Mpc) and large scales, whereas ξ and ζ estimators based on the idea of equation (2.116) have severe signal to noise problems on large scales. It is useful to define, based on (2.150), a reduced bispectrum from the ratio

$$Q(k_1, k_2, k_3) = \frac{B_{123}}{P_1 P_2 + P_2 P_3 + P_3 P_1} \tag{2.151}$$

This proves useful in that several orders of magnitude in B can be reduced to Q variations on the order of 1. Furthermore, perturbation theory predictions for Q reduce to (2.150) in a special case ($P \sim k^n$, $k_1 = k_2 = k_3$), and in general the Q functional dependence from perturbation theory pertains to large-scale measurements which confirms that gravitational instability is in effect.

2.3.3 Primordial Fluctuations, Galaxies, and Bias

It is conventional to assume that the primordial density fluctuations are a homogeneous Gaussian random field. Such a field is completely specified by its 2-point correlations, as the connected N-point correlation functions are all zero for $N > 2$. One can argue, roughly speaking, that the validity of this assumption is a natural consequence of early fluctuations representing noise with random phases in Fourier space, invoking the central limit theorem (Zel'dovich & Novikov 1983).

With the Gaussian random field assumption, it follows, for example, that the one-point probability distribution $\mathcal{P}$ for the initial $\delta(\mathbf{x})$ field is Gaussian

$$\begin{aligned} \mathcal{P}(\delta)\,d\delta &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\delta^2}{2\sigma^2}\right) d\delta \\ \sigma^2 &= \xi(0) = \langle \delta(\mathbf{x})\delta(\mathbf{x}) \rangle \end{aligned} \tag{2.152}$$

where $\sigma \ll 1$ is the same at all points. In fact, the random field description involves a probability functional $\mathcal{P}[\delta(\mathbf{x})]$, and its related moment generating functional, as explained by Fry (1997), Borgani (1995), and Bertschinger (1992). The homogeneity and Gaussianity assumptions simplify matters greatly. We rely heavily on these simplifications especially in the numerical simulations we present below Thus we have that the N-point probability distribution for δ_1 δ_N is Gaussian with covariance function $\xi(r_{ij})$ with all higher connected moments vanishing—in analogy with the single variable Gaussian distribution In particular the covariance function is diagonal in k space and so the components of $\delta(\mathbf{k})$ are distributed normally

$$\begin{aligned}
\mathcal{P}(\tilde{\delta}) &= \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\Big(-\frac{\tilde{\delta}^2}{2\sigma_k^2}\Big)\, d\tilde{\delta} \\
\sigma_k^2 &= P(k)/2
\end{aligned} \qquad (2\ 153)$$

where $\tilde{\delta}$ stands for both the real and imaginary parts of $\delta(\mathbf{k})$ The power per mode is denoted by $P(k)$ as in equation (2 142) In terms of modulus and phase this says that the phases are uniformly random and that the modulus $|\delta(\mathbf{k})|$ is Rayleigh distributed

$$\begin{aligned}
\delta(\mathbf{k}) &= |\delta(\mathbf{k})|\, e^{i\theta} \\
\mathcal{P}(|\tilde{\delta}(\mathbf{k})|)\, d|\delta(\mathbf{k})| &= \frac{|\delta(\mathbf{k})|}{P(k)} \exp\Big(-\frac{|\delta(\mathbf{k})|^2}{2P(k)}\Big)\, d|\delta(\mathbf{k})| \\
\mathcal{P}(\theta)\, d\theta &= \frac{1}{2\pi}\, d\theta
\end{aligned} \qquad (2\ 154)$$

where the mean and standard deviation of the $|\delta(\mathbf{k})|$ distribution are

$$\begin{aligned}
\langle|\delta(\mathbf{k})|\rangle &= \sqrt{\frac{\pi}{2} P(k)} \\
\sigma_{|\delta(\mathbf{k})|} &= \sqrt{\frac{4-\pi}{2} P(k)}
\end{aligned} \qquad (2\ 155)$$

The Rayleigh distribution (2 154) is obtained by treating the random variable $|\delta(\mathbf{k})|$ as derived from the independent real and imaginary parts of $\delta(\mathbf{k})$ Similarly the distribution for $|\tilde{\delta}(\mathbf{k})|^2$ is exponential

$$\mathcal{P}(|\delta(\mathbf{k})|^2)\, d|\delta(\mathbf{k})|^2 = \frac{1}{P(k)} \exp\Big(-\frac{|\delta(\mathbf{k})|^2}{P(k)}\Big)\, d|\delta(\mathbf{k})|^2 \qquad (2\ 156)$$

with mean and standard deviation given by

$$\begin{aligned}
\langle|\delta(\mathbf{k})|^2\rangle &= P(k) \\
\sigma_{|\delta(\mathbf{k})|^2} &= \sqrt{2} P(k)
\end{aligned} \qquad (2\ 157)$$

Primordial power is conveniently parametrized as a featureless power law $P(k) \propto k^n$. The Zel'dovich scale-free spectrum (Harrison 1970; Peebles & Yu 1970; Zel'dovich 1972), which has spectral index $n = 1$, is favored theoretically. Also, inflationary theories predict $P \sim k$ fairly robustly. Using the COBE measurements to tie down the large-scale fluctuation spectrum indicates that $n = 1$ is likely (Gorski et al. 1994).

The Zel'dovich spectrum is scale-free in the sense that gravitational potential fluctuations have equal amplitude on all scales (Bertschinger 1993). The potential fluctuations on a physical scale k^{-1} are $\sim k^3|\tilde{\varphi}|^2$. The scale-free condition requires a potential spectrum $|\tilde{\varphi}|^2 \sim k^{-3}$, but $\varphi \sim \delta/k^2$ from the Poisson equation, and thus we need $|\delta|^2 \sim k$. In the language of density fluctuations, the interpretation of $P(k) \sim k$ as scale-free refers to the following fact. A mean square fluctuation $\sim k^3 P(k)$ of physical scale k^{-1}, growing according to superhorizon perturbation theory, has a constant ($\approx 10^{-5}$) amplitude as the wavelength enters the horizon ($d_H \sim t$) for all scales.

While fluctuations are evolving early in the linear regime, with $z > z_{eq}$ (the matter radiation equality epoch is $z_{eq} \approx 10^4$), physical processes modified the primordial spectrum prior to the onset of structure formation. This is codified in a linear transfer function T which describes evolution from some primordial time t_p to some later time t_i, which acts as the initial time for the dynamical evolution in our equations of motion. Recalling that $\delta \sim A(t)$ in linear theory, we connect the early power spectrum $P(k, t_p)$ to the spectrum at t_i, corresponding to $z \approx 1000$, with T^2:

$$P(k, t_i) = \left[\frac{A(t_i)}{A(t_p)}\right]^2 T^2(k, t_i) P(k, t_p) \tag{2.156}$$

For physical wavelengths $\gg t_{eq}$, the horizon at z_{eq}, the slope of $P(k)$ (at small k) is unaffected. The shape of $P(k)$ for lengths $k^{-1} \lesssim t_{eq}$ (large k) is affected, and the details depend on the particular model being used. Transfer functions for various models are reviewed by Efstathiou (1990). A convenient numerical fit from the original CDM paper (Peebles 1982) is

$$T^2 = (1 + \alpha k + \beta k^2)^{-2} \tag{2.159}$$

where the numerical factors are provided by Peebles (1982). There was an onslaught of CDM and, for a while, HDM papers in the 1980's. Some numerical fits to the transfer functions for such models are provided in Bardeen et al. (1986).

The CDM function in (2.158) shows the typical bend

$$\begin{aligned} P &\sim k \quad k \ll k_{eq} \\ P &\sim k^{-3} \quad k \gg k_{eq} \end{aligned} \tag{2.160}$$

which shows the preferred length scale $k_{eq}^{-1} \approx 10/\Omega_0 h^2$ Mpc, the horizon at matter-radiation equality, though the transition from k to k^{-3} is gradual. This transition from a primordial $P \sim k^n$ to $P \sim k^{n-4}$ follows from the fact that fluctuations of non-relativistic (cold) matter which enter the horizon at $z > z_{eq}$ practically do not grow until $z \gtrsim z_{eq}$, since the subhorizon perturbations in the dominant radiation oscillate like an acoustic wave (Peebles 1980).

The HDM transfer function (Bond & Szalay 1983) is

$$T \approx 10^{-(k/k_\nu)^{1.5}} \tag{2.161}$$

corresponding to a universe dominated by collisionless neutrinos. The free streaming damping scale is $13(\Omega_\nu h^2)^{-1}$ Mpc which is ≈ 40 Mpc for a 30 eV neutrino. Small scales are washed out, and large structures form first in what would be the most direct application of Zel'dovich pancake formation. There are problems in timing (galaxies would form too late) and $\Delta T/T$ (cosmic background radiation fluctuations would need to be too large) with the simplest HDM models.

The initial linear regime $P(k)$ can be normalized in various ways, as reviewed by Bertschinger (1992). Two ways are provided by Peebles (1982). Large-scale normalization (at scales $\approx 10^4\,h^{-1}$ Mpc) is based on matching the small $\langle\delta^2\rangle_H$ fluctuations at the horizon with COBE quadrupole anisotropies. Matching σ_8^2, defined to be the $\langle\delta^2\rangle_R$ at $R = 8\,h^{-1}$ Mpc, with $\sigma_8 \approx 1$ from observations is a smaller scale normalization; this method uses equation (2.146). These two normalizations are equivalent if the linear $P(k)$ accurately spans such scales.

The last point which we mention here concerns galaxies, which form the observational basis of our subject. We must assume—rather cavalierly perhaps, because we know it is an oversimplification—that the present observable distribution of galaxies is directly related to the random field $\delta(\mathbf{x})$. Discreteness of the sample complicates matters, but this is not a fundamental problem for the model (homogeneous random process) that we use. The problem, which we mentioned at the outset, is that so much physics (galaxy formation, galaxy statistics) must be ignored in the assumption. Assuming that we have a handle on $\delta(\mathbf{x})$, we may use the equations of motion developed above to evolve the density (and velocity and potential) forward or backward in time.

That said, there is one last complication which we may at least try to model simply. This concerns the relation between $\delta(\mathbf{x})$ and $\delta_g(\mathbf{x})$, the matter and galaxy distributions. If there is no relation, then physicists would need to go back to the drawing board. Possibly, galaxies faithfully trace the mass, so that a smoothed δ_g field would really correspond to δ. At the least, there should be some functional relationship, and if the range of influence is not large, then we would have a local bias (Coles 1992, Kaiser 1984), with $\delta_g(\mathbf{x}) = f(\delta(\mathbf{x}))$. The simplest case is linear bias, where fluctuations are linearly stretched, $\delta_g = b_1\delta$.

2.4 Perturbation Theory for Statistics

We now turn to some of the higher order perturbation theory results. The linear solutions introduced in §2.2.4 are of obvious importance, as they represent the leading order term, which is the largest when $\langle \delta^2 \rangle \ll 1$. As such they describe the initial deviations from uniformity.

Going to second or third order gives results which are surprisingly useful in analysing both N-body simulations and actual galaxy survey data. This might be surprising because the higher order solutions are describing the development of nonlinearities. The linear density solution,

$$\begin{aligned} \delta^{(1)}(\mathbf{x}, t) &= A(t)\, \delta_0(\mathbf{x}) \\ \delta_0(\mathbf{x}) &= \delta(\mathbf{x}, t_i) \end{aligned} \qquad (2.162)$$

is modified by the addition of higher order terms in the perturbation series, e.g. of order $\sim (\delta^{(1)})^2$, which are nonlocal and naturally change the shape of the initial δ_0. But as the higher order terms become important, as $\delta \approx 1$, all the higher order terms are large, and so the usefulness of keeping just the second and third order terms is not obvious a priori. Finally, the convergence of a perturbation series expansion has not been proven. Despite those grievances, it is fair to say that higher order perturbation theory is quite relevant in what one might call the quasi-linear regime (Fry 1984, Jain & Bertschinger 1994).

We utilize the formalism and notation developed by Peebles (1980), Fry (1984) and Goroff et al. (1986). This is sometimes referred to as Eulerian perturbation theory since we use comoving spatial coordinates. Thus the distinction is made between this and Lagrangian perturbation theory, which uses comoving material coordinates (which follow particle trajectories).

The fundamental equation, developed in §2.3.3, is

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} - 4\pi G\rho\delta = 4\pi G\rho\delta^2 + \frac{1}{a^2}\nabla\varphi \cdot \nabla\delta + \frac{1}{a^2}\nabla_i\nabla_j[(1+\delta)v^i v^j] \qquad (2.163)$$

ignoring the entire right hand side, which contains higher order terms, we get $\delta^{(1)}$ as in (2.162), the (rescaled) gravitational potential which is the solution to $\nabla^2\Delta^{(1)} = \delta^{(1)}$, and the velocity $\mathbf{v}^{(1)} = -a\dot{A}/A\nabla\Delta^{(1)}$.

We write the perturbation series as

$$\delta = \delta^{(1)} + \delta^{(2)} + \delta^{(3)} + \dots \tag{2.164}$$

where the higher terms are of higher order in smallness (for $\delta \ll 1$). In particular, the iterative solutions show that $\delta^{(n)} \sim A^n \mathcal{O}(\delta_0^n)$. For $\Omega = 1$, using the growing mode $A \sim t^{2/3}$, the solutions are exactly of the form

$$\delta(\mathbf{x}, t) = \sum_{n=1}^{\infty} A^n(t)\delta_n(\mathbf{x}) \tag{2.165}$$

as shown by Goroff et al. (1986). Results were obtained by Bouchet et al. (1992) for arbitrary Ω, and for $\Omega \approx 1$ the series is nearly of the form (2.165), with very weak Ω dependence.

To obtain the differential equation for $\delta^{(2)}$, we use the linear results, in terms of $\delta^{(1)}$ and $\Delta^{(1)}$, on the right hand side of equation (2.163), keeping terms of order $(\delta^{(1)})^2$; on the left hand side we use $\delta^{(2)}$, as per a standard iterative scheme. In the $\Omega = 1$ case, this leads to the following second order solution

$$\delta^{(2)} = \frac{5}{7}(\delta^{(1)})^2 + \delta^{(1)}_{,i}\Delta^{(1)}_{,i} + \frac{2}{7}\Delta^{(1)}_{,ij}\Delta^{(1)}_{,ij} \tag{2.166}$$

where $f_{,i} \equiv \partial f/\partial x^i$. This follows the time dependence as shown in (2.165) and also has $\langle\delta^{(2)}\rangle = 0$, as expected.

As an illustrative application, we may compute the evolution of the third moment $\langle\delta(\mathbf{x}, t)^3\rangle$ from Gaussian initial conditions (where all connected moments of δ_0 above the 2 point function vanish) We assume that $\Omega = 1$ To leading nonvanishing order we have (recall that $\delta^{(1)}$ is a Gaussian field whose third moment vanishes)

$$\begin{aligned}
\langle\delta^3\rangle &= \langle[\delta^{(1)} + \delta^{(2)} + \dots]^3\rangle \\
&\approx 3\langle\delta^{(1)2}\delta^{(2)}\rangle \\
&= \frac{34}{7}\langle\delta^{(1)2}\rangle^2
\end{aligned} \tag{2 167}$$

This result based on (2 166) is shown by Peebles (1980) Details are also provided by Fry and Scherrer (1993), who generalize to a case with non-Gaussian initial conditions The local skewness parameter $S_3 \equiv \langle\delta^3\rangle/\langle\delta^2\rangle^2$ is thus independent of time and the form of the initial 2-point correlations If we consider the initially Gaussian distribution function for the δ values we may ask how the skewness (in the usual sense of the word) of the distribution evolves This normalized skewness is

$$\langle\delta^3\rangle/\langle\delta^2\rangle^{3/2} = \frac{34}{7}\langle\delta^2\rangle^{1/2} \tag{2 168}$$

and it increases with time as $A(t)$, to lowest order

Further hierarchical amplitudes S_N coming from the connected N-point correlation function average $\langle\delta^N\rangle_c$ can be similarly defined $\langle\delta^N\rangle_c = S_N\langle\delta^2\rangle^{N-1}$ and the S_N are constant in the Einstein–de-Sitter case, as in (2 167) and largely independent of cosmological parameters in general (Goroff et al 1986 Bernardeau 1992) For a possible relation to observation we need to use the smoothed density field δ_W, as defined in equation (2 144) In practice the moments of counts in cells as in equations (2 123–2 126), are related to volume averaged correlation functions (apart from discreteness corrections) and the hierarchical scaling parametrized by S_N is an observable effect Smoothing at scale R the third moment of δ_W gives

$$\begin{aligned}
S_3 &= \frac{34}{7} + \gamma_1 \\
\gamma_1 &= \frac{d\log\langle\delta^2\rangle_R}{d\log R}
\end{aligned} \tag{2 169}$$

where γ_1 is scale dependent The γ_1 term depends on the shape of $P(k)$ and the shape of the filter in general For initial power $P \sim k^n$ with a top hat filter as in (2.145) we have $\langle\delta^2\rangle_R \sim R^{-(n+3)}$ which gives the rather nice result that $S_3 = 34/7 - (n+3)$ Bernardeau (1994) includes this and higher order results

We focus on the situation in Fourier transform space as introduced in §2.3.2 where comoving coordinates $\mathbf{x}$ and wavevectors $\mathbf{k}$ describe the complimentary domains Linear perturbation theory viewed in the Fourier domain says that fluctuations at different wavelengths evolve independently Nonlinear nonlocal effects—represented to lowest nonvanishing order by equation (2.166)—are seen as mode coupling in k-space

Consider the full $\delta(\mathbf{x}, t)$ equation (2.163) viewed in Fourier space Considering the lowest order nonlinear effects of order $\delta^{(1)2}$ we would have an equation of the following form for $\delta(\mathbf{k}\ t) \equiv \delta_{\mathbf{k}}$,

$$\frac{\partial^2 \delta_{\mathbf{k}}}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta_{\mathbf{k}}}{\partial t} - 4\pi G\bar{\rho}\delta_{\mathbf{k}} = \int d^3k' d^3k'' \delta_{\mathbf{k}'}\delta_{\mathbf{k}''} F(\mathbf{k}, \mathbf{k}', \mathbf{k}'')\delta_D(\mathbf{k} - \mathbf{k}' - \mathbf{k}'') \quad (2.170)$$

For example, this generates harmonics for $\mathbf{k}$ values which are not even present in the initial $P(k)$ In general the evolution of a mode $\mathbf{k}$ is determined by coupling of all modes $\mathbf{k}'$, $\mathbf{k}''$ whose sum is $\mathbf{k}$ (required by spatial homogeneity) Equation (2.170) represents perturbations up to second order in that $\delta^{(1)}(\mathbf{k}\ t)$ is the solution to the equation with the right hand side zero, and $\delta^{(2)}_{\mathbf{k}}$ comes from two interacting modes $\delta^{(1)}_{\mathbf{k}'}\delta^{(1)}_{\mathbf{k}''}$

In particular for $\Omega = 1$ with $t^{-2} = 6\pi G\bar{\rho}$ and $a \sim A \sim t^{2/3}$ we keep track of the lowest nonlinear terms to get

$$\begin{aligned} &\frac{\partial^2 \delta_{\mathbf{k}}}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta_{\mathbf{k}}}{\partial t} - 4\pi G\bar{\rho}\delta_{\mathbf{k}} = \\ &\frac{28G\bar{\rho}}{3}\int \frac{V dk_1^3}{(2\pi)^3}\left[\frac{5}{7} + \frac{\mathbf{k}_1 \cdot (\mathbf{k} - \mathbf{k}_1)}{|\mathbf{k} - \mathbf{k}_1|^2} + \frac{2}{7}\left(\frac{\mathbf{k}_1 \cdot (\mathbf{k} - \mathbf{k}_1)}{k_1|\mathbf{k} - \mathbf{k}_1|}\right)^2\right]\bar{\delta}_{\mathbf{k}_1}\delta_{\mathbf{k}-\mathbf{k}_1} \end{aligned} \quad (2.171)$$

Neglecting the right hand side of (2.171) gives $\delta^{(1)} \sim A(t)$, the transform of the linear solution. Including the right hand side, and equating terms of second order, gives $\delta^{(2)}_{\mathbf{k}} \sim A^2(t)$, the transform of the x space second order solution (2.166), which we write as

$$\delta^{(2)}_{\mathbf{k}} = \int \frac{V dk_1^3}{(2\pi)^3}\frac{V dk_2^3}{(2\pi)^3} G_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2)\, \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \,[\delta_D] \tag{2.172}$$

$$G_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) = \frac{5}{7} + \frac{1}{2}\frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{2}{7}\left(\frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2}\right)^2 \tag{2.173}$$

$$[\delta_D] \equiv \left[\frac{(2\pi)^3}{V}\delta_D(\mathbf{k} - \sum \mathbf{k}_i)\right] \tag{2.174}$$

The relations among δ, $\mathbf{v}$, and Δ in k-space are quite convenient, with differentiation $\nabla_{\mathbf{x}}$ bringing about factors of $\mathbf{k}$. But products in x space are of course transformed into convolutions. The tradeoff works in favor of transform space, however, as shown by Fry (1984) and Goroff et al. (1986) when third and higher order perturbations are considered. This advantage in computational formalism is in addition to the advantages of Fourier spectra as regards to analysis of large-scale data.

The mode coupling kernel $G_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2)$ is the first in a sequence which specifies the perturbation series (2.154) in Fourier space. Even in unsymmetrized form, the third order kernel (Fry 1984) has a lengthy formula not reproduced here. The (nonsymmetrized) fourth order kernel takes a full page to write down, and Goroff et al. (1986) supplied this as well as a recursion for $G_n(\mathbf{k}_1, \ldots, \mathbf{k}_n)$ and an analogous kernel for the velocity divergence (which specifies $\mathbf{v}$ since it is a gradient). Jain and Bertschinger (1994) clarified the formalism and compared second order power spectrum calculations (which go to third order in perturbation theory) to N-body simulations. Fry (1994a) placed the higher order power spectrum calculations in compact form and produced a calculation up to fifth order in perturbation theory for the case of a top-hat power spectrum. This clarified how small k and large k power is generated from an initial spectrum which is nonzero for only a finite range in k.

Using the second order kernel (2.173), we may calculate the bispectrum to lowest nonvanishing order from

$$B(k_1, k_2, k_3) = \langle \delta^{(1)}_{\mathbf{k}_1} \delta^{(1)}_{\mathbf{k}_2} \delta^{(2)}_{\mathbf{k}_3} \rangle + (\text{cyc.}) \tag{2.175}$$

Recall that $\sum \mathbf{k}_i = 0$: the triangle must be closed on the k-space grid. We find that

$$\begin{aligned} B(k_1, k_2, k_3) &= Q_{12} P_1 P_2 + Q_{23} P_2 P_3 + Q_{31} P_3 P_1 \\ Q_{ij} &= \frac{10}{7} + \mathbf{k}_i \cdot \mathbf{k}_j \left(\frac{k_i}{k_j} + \frac{k_j}{k_i} \right) + \frac{4}{7} (\mathbf{k}_i \cdot \mathbf{k}_j)^2 \end{aligned} \tag{2.176}$$

Thus we have a perturbative result for $Q \equiv B_{123}/(P_1P_2 + P_2P_3 + P_3P_1)$, and the shape dependence of Q provides a signature for gravitational instability. There are two particularly simple, useful triangle configurations (they must be closed triangles by homogeneity). For equilateral triangles of side k ($\mathbf{k}_i \cdot \mathbf{k}_j = -\frac{1}{2}$), we have $Q_{eq}(k)$ as a function of scale. If we fix k_1 and k_2 ($k_1 = 2k_2$ proves to be a practical choice) and let $\cos\theta = \mathbf{k}_1 \cdot \mathbf{k}_2$, then $Q(\theta)$ is a useful parametrization based on shape, and it spans scales from $|\mathbf{k}_1 - \mathbf{k}_2|$ down to $|\mathbf{k}_1 + \mathbf{k}_2|$. Fry (1994b) has used the $Q(\theta)$ statistic as a useful method to isolate the effects of biasing from clustering and parameters such as Ω_0 and H_0.

CHAPTER 3
ZEL'DOVICH APPROXIMATION

3.1 The Equations of Motion and the Approximation

3.1.1 Material Coordinate Formulation

The Zel'dovich approximation of gravitational instability takes place within a Lagrangian, or material coordinate, formulation of the cosmological fluid equations (Zel'dovich 1970). In the last chapter, we introduced the conventional Eulerian, or spatial coordinate, approach to the equations and the resulting perturbation theory. As in classical, earthly fluid mechanics, both approaches should be equivalent yet complimentary, each having its own advantages and disadvantages (Truesdell 1954).

We consider the paths of individual particles or fluid elements beginning at a time t_i (the decoupling era, for example) when the universe was highly uniform. The comoving coordinates at time t_i are taken as labels, or material coordinates, which identify the particles throughout their trajectories. We denote these material coordinates by $\mathbf{q}$. The comoving coordinates $\mathbf{x}$ following a trajectory are $\mathbf{x}(\mathbf{q}, t)$. Formally, for convenience, we take $t_i = 0$, so we have $\mathbf{x}(\mathbf{q}, 0) = \mathbf{q}$. The trajectories in proper coordinates are given by $\mathbf{r}(\mathbf{q}, t)$, which includes both the Hubble flow and peculiar motions. This method is naturally discrete (though it need not be: $\mathbf{q}$ can be a continuous label) and as such is ready-made for numerical simulations. In this picture fluid motion is viewed as the Lagrangian map $\mathbf{q} \mapsto \mathbf{x}(\mathbf{q}, t)$, most conveniently studied in comoving coordinates $\mathbf{x}$ (as we have specified) rather than in the $\mathbf{r}$ coordinates

We write this $\mathbf{x}(\mathbf{q}, t)$ mapping as

$$\mathbf{x} = \mathbf{q} + \Psi(\mathbf{q}, t) \tag{3.1}$$

where the displacement $\Psi(\mathbf{q}, t)$ describes the gravitational instability, so $\Psi \equiv 0$ represents a pure Hubble flow.

In parallel with the development in Chapter 2, we may take an ideal fluid (collisional) or a collisionless viewpoint. Taking a conventional ideal fluid continuum approach, each fluid element will be described by its mass, position, and velocity. Furthermore, each particle possesses thermodynamic properties, and the pressure forces must be accounted for. We follow the collisionless approach, which again is most appropriate for the era of structure formation. Therefore we consider a collisionless gas of particles which interact by gravitation only. The problem is thus simplified to gravitational instability only, as particles are endowed only with mass, position and velocity. The particles follow geodesics of the perturbed Robertson Walker metric (2.72). The equations of motion (Newtonian limit) are those we developed in §2.2.3, and again we utilize the separation of the Hubble flow from the the overall trajectories, with the dynamics following from the Lagrangian

$$L = \frac{1}{2} m a^2 \mathbf{x}^2 - m\varphi \tag{3.2}$$

where m is a particle mass, $a(t)$ is the expansion factor, $\mathbf{x}$ is the comoving coordinate and φ is the peculiar gravitational potential.

The equations of motion in comoving coordinates are

$$\mathbf{v} = a\frac{d\mathbf{x}}{dt} \tag{3.3}$$

$$\frac{d\mathbf{v}}{dt} + \frac{a}{a}\mathbf{v} = -\frac{1}{a}\nabla\varphi \tag{3.4}$$

The time derivative d/dt is convective (evaluated at fixed $\mathbf{q}$)

To this we must add the Poisson equation

$$\nabla^2\varphi(\mathbf{x},t) = 4\pi G\bar{\rho}a^2\delta(\mathbf{x},t) \tag{3.5}$$

Expressed in terms of the peculiar acceleration $\boldsymbol{g}$ we have

$$\begin{gathered} \boldsymbol{g} = -\frac{1}{a}\nabla\varphi \\ \nabla\cdot\boldsymbol{g} = -4\pi G\bar{\rho}a\delta \quad \nabla\times\boldsymbol{g} = 0 \end{gathered} \tag{3.6}$$

Note that the Poisson equation requires the density contrast field in spatial coordinates $\mathbf{x}$

In the Lagrangian formulation, continuity is assured by simply applying conservation of mass to the trajectories, and before multistreaming occurs in the $\mathbf{x}(\mathbf{q},t)$ mapping this is a straightforward procedure We equate the mass in the fluid element at $\mathbf{q}$ at time t_i which is $\bar{\rho}(t_i)a^3(t_i)d^3q$ to the mass at $\mathbf{x}(\mathbf{q},t)$ which is $\rho(\mathbf{x},t)a^3(t)d^3x$, and note that $\bar{\rho}a^3$ is constant in time Therefore we write

$$\rho(\mathbf{x},t)d^3x = \bar{\rho}(t)d^3q\,, \tag{3.7}$$

noting that this applies exactly for single stream flow The deformation of the volume element is described by the Jacobian of the mapping,

$$d^3x = \left\|\frac{\partial\mathbf{x}}{\partial\mathbf{q}}\right\| d^3q\,, \tag{3.8}$$

where $\|\ \|$ denotes the determinant and so we define the Jacobian $J(\mathbf{q},t)$ and a deformation tensor J_{ij} as

$$J_{ij} = \frac{\partial x^i}{\partial q^j} \tag{3.9}$$

$$J = \|J_{ij}\| \tag{3.10}$$

In terms of the density contrast δ defined by $\rho = \bar{\rho}(1+\delta)$ continuity is expressed as

$$1 + \delta = J^{-1} \tag{3.11}$$

The peculiar velocity $\mathbf{v}[\mathbf{x}(\mathbf{q},t),t]$ can be written as

$$\mathbf{v} = a\frac{d\mathbf{x}}{dt} = a\Psi \tag{3.12}$$

where the d/dt represents the convective $(\partial/\partial t)_{\mathbf{q}}$. In general, the $\mathbf{q}$ and $\mathbf{x}$ derivatives are related by

$$\frac{\partial}{\partial x^i} = \frac{\partial q^j}{\partial x^i}\frac{\partial}{\partial q^j} = J^{-1}_{ij}\frac{\partial}{\partial q^j} \tag{3.13}$$

$$\left(\frac{\partial}{\partial t}\right)_{\mathbf{q}} = \left(\frac{\partial}{\partial t}\right)_{\mathbf{x}} + \frac{1}{a}(\mathbf{v}\cdot\nabla_{\mathbf{x}})_t \tag{3.14}$$

The Eulerian continuity equation is

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot(1+\delta)\mathbf{v} = 0 \tag{3.15}$$

$$\frac{d\delta}{dt} + \frac{1}{a}(1+\delta)\nabla\cdot\mathbf{v} = 0 \tag{3.16}$$

and we have rewritten (3.15) using $(\partial\delta/\partial t)_{\mathbf{q}}$.

One of the secrets to the success of the Lagrangian approach and its associated perturbation theory is that the Eulerian continuity equation is satisfied automatically. Note that δ does not appear as a dynamic variable in the equations of motion (3.3) and (3.4), while it of course must appear in the Eulerian continuity equation. It is easy to show how this works out. We compute the convective time derivative of δ,

$$\left(\frac{\partial\delta}{\partial t}\right)_{\mathbf{q}} = \frac{\partial}{\partial t}J^{-1} = (1+\delta)\frac{1}{J}\frac{\partial J}{\partial t} \tag{3.17}$$

Expanding the determinant and taking its derivative, we obtain

$$J(\mathbf{q},t) = \frac{1}{3!}\epsilon_{ijk}\epsilon_{lmn}\frac{\partial x^l}{\partial q^i}\frac{\partial x^m}{\partial q^j}\frac{\partial x^n}{\partial q^k} = \epsilon_{ijk}\frac{\partial x^1}{\partial q^i}\frac{\partial x^2}{\partial q^j}\frac{\partial x^3}{\partial q^k} \tag{3.18}$$

$$\frac{\partial J}{\partial t} = \epsilon_{ijk}\left\{\frac{\partial}{\partial t}\frac{\partial x^1}{\partial q^i}\right\}\frac{\partial x^2}{\partial q^j}\frac{\partial x^3}{\partial q^k} + [\text{2 cyclic terms}] \tag{3.19}$$

where ϵ_{ijk} is the totally antisymmetric permutation tensor. The $\{\ \ \}$ term is

$$\frac{\partial}{\partial t}\frac{\partial x^1}{\partial q^i} = \frac{\partial\Psi^1}{\partial q^i} = \frac{1}{a}\frac{\partial v^1}{\partial x^m}\frac{\partial x^m}{\partial q^i} \tag{3.20}$$

and so only the $m = 1$ term survives to give $a^{-1}\partial v^1/\partial x^1 J$ as the first of the three terms in (3.19). Therefore we have

$$\frac{1}{J}\frac{\partial J}{\partial t} = \frac{1}{a}\nabla \cdot \mathbf{v} \tag{3.21}$$

which establishes the consistency of the two approaches, comparing (3.16) with (3.17). This all follows, more concisely, from the identity

$$\frac{\partial}{\partial t}\log \|J_{ij}\| = J^{-1}_{ij}\frac{\partial J_{ji}}{\partial t} \tag{3.22}$$

The Eulerian velocity field due to gravitational instability is irrotational. Recall that rotational modes for $\mathbf{v}$ do not couple to gravity, and from the Kelvin circulation theorem in an expanding universe, such modes decay as a^{-1}. The displacement mapping $\Psi(\mathbf{q}, t)$ could, in principle, encompass such transverse modes in general and in perturbation theory (Buchert 1992). As appropriate to large-scale structure formation, we have

$$\nabla \times \mathbf{v} = 0 \tag{3.23}$$

in the Eulerian picture. Using (3.13) and (3.12), this important irrotational condition translates to

$$\epsilon_{ijk} J^{-1}_{lj} J_{kl} = 0 \tag{3.24}$$

in material coordinates, and this is valid during single stream flow and, later on, outside of multistreaming regions which occur as pancakes, filaments, and clumps are formed. Expressing the equation of motion following a particle, by taking the divergence of (3.4) and utilizing (3.13), we have

$$J^{-1}_{ji}\frac{\partial}{\partial t}\left(a^2\frac{\partial J_{ij}}{\partial t}\right) = -4\pi G a^2 (J^{-1} - 1) \tag{3.25}$$

which, together with (3.24), provides the dynamics of the $\Psi(\mathbf{q}, t)$ map.

This reduction of the equations of motion to (3.24) and (3.25) is similar to what transpires in Eulerian coordinates when deriving the Bernoulli equation. For potential flow, we may integrate the Euler equation to obtain the Bernoulli equation in a particularly simple form. Writing $\mathbf{v}$ as a gradient, $\mathbf{v} = -\nabla\varphi_v/a$, the Euler equation

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{a}\mathbf{v}\cdot\nabla\mathbf{v} + \frac{\dot{a}}{a}\mathbf{v} = -\frac{1}{a}\nabla\varphi \tag{3.26}$$

reduces to

$$\frac{\partial \varphi_v}{\partial t} - \frac{1}{2a^2}(\nabla\varphi_v)^2 = \varphi \tag{3.27}$$

3.1.2 The Zel'dovich Approximation

Despite the complexity of the large-scale structure problem overall, there is an underlying simplicity in our restricted problem which was recognized early on by the famed Russian physicist Yakov Borisovich Zel'dovich (1914 – 1987). In 1970, Ya. B. Zel'dovich published the article "Gravitational Instability: An Approximate Theory for Large Density Perturbations" (Zel'dovich 1970). The paper is brief, intuitively and physically motivated, and (as came to be realized some 10 to 15 years later) important. Recently, it has loomed even more significant as the approximation introduced therein has been found to be much more general than first suspected. The matter of the universe may be primarily baryonic or nonbaryonic; the nonbaryonic component may be "hot" or "cold". Regardless, the approximation has something to say about structure on large scales.

The linear theory of perturbations, which we addressed in §2.2.4, is well established. It can be thought of as the exact theory in the realm of small perturbations $\langle\delta^2\rangle \ll 1$. As we discussed in Chapter 2, when the perturbation series is continued to second and third order (and beyond) the N-point correlation hierarchy may be calculated (Peebles 1980; Fry 1984; Goroff et al. 1986) for $n = 3$, $n = 4$, etc. in hopes that the observed galaxy correlations may retain a statistical imprint of the perturbation theory results. Naively, one would expect the perturbation series results to have a bearing only on the largest scales (where $\langle\delta^2\rangle$ is still small) but there is evidence that this imprint is discernible even in the nonlinear regime Nonetheless perturbation theory is generally limited by the constraint $\langle\delta^2\rangle \lesssim 1$ This is merely to suggest that an extrapolation of perturbation theory may be in order This is somewhat obvious However when Zel'dovich made his suggestion to do so, there was no observational evidence to speak of

The basic idea of the approximation—remarkably simple in hindsight—is to extrapolate linear gravitational theory into the nonlinear regime by following perturbations of particle trajectories, rather than tracking perturbations in the density field As such, the Lagrangian formalism which has some inherent advantages in analyzing cosmological perturbations was introduced by Zel'dovich The approximation involves separating the space and time dependence in the Lagrangian mapping, by writing $\Psi(\mathbf{q}, t) = b(t)\mathbf{u}(\mathbf{q})$, and following the particle initially at $\mathbf{q}$ to its Eulerian comoving coordinate $\mathbf{x}$, related to the proper the Eulerian proper coordinates by $\mathbf{r} = a\mathbf{x}$. Thus the Zel'dovich approximation has the following simple form,

$$\mathbf{r}(\mathbf{q}, t) = a(t)\left[\mathbf{q} + b(t)\mathbf{u}(\mathbf{q})\right], \tag{3.28}$$

where the growing function $b(t)$ characterizes the growth of the unstable gravitational fluctuations It turns out that $b(t)$ is just the growing mode of linear fluctuations $A(t)$, though we keep the $b(t)$ notation of Shandarin and Zel'dovich (1989)

Linear perturbation theory in the Eulerian picture states that

$$\delta(\mathbf{x}, t) = A(t)\delta_0(\mathbf{x}) \tag{3.29}$$

The function $A(t)$ is the growing solution of (2.97) and the primal density contrast $\delta_0(\mathbf{x})$ thus retains its shape in the initial stages of growth The peculiar velocity field is

$$\mathbf{v}(\mathbf{x}, t) = -a\frac{A}{A}\nabla\Delta \tag{3.30}$$

where Δ, the rescaled first order peculiar gravitational potential

$$\Delta = \frac{1}{4\pi G\rho a^2}\varphi \tag{3.31}$$

is a solution to the Poisson equation, which to this order reads

$$\nabla^2\Delta = A(t)\delta_0(\mathbf{x}) \tag{3.32}$$

The peculiar acceleration $g(\mathbf{x}, t)$ is parallel to $\mathbf{v}(\mathbf{x}, t)$ and, of course, $\nabla \times \mathbf{v} = 0$. For the time dependence in linear theory, we may write $\delta \sim A$. The potential is nearly time independent, $\varphi \sim A/a$, and for $\Omega = 1$ we have $A \sim a$ and φ is constant in time. The rescaled potential obeys $\Delta \sim A$, and the peculiar velocity follows $\mathbf{v} \sim a\dot{A}$.

To extend (3.29) as it stands leads to problems. The initial power spectrum $\langle|\delta_0(\mathbf{k})|^2\rangle$ evolves by a factor of $A^2(t)$ while retaining its shape, and we know this result, while valid early on and useful for very large scales, is not reliable as a detailed extrapolation, since mode coupling is ignored. Moreover, the following inconsistency is most egregious: while $A(t)$ grows, (3.29) would say that in some regions δ might have a value of, say, 1 or 2, while in other regions δ acquires a meaningless value less than -1.

That a naive extension of linear theory goes astray is no surprise. During linearization, the $\nabla \cdot (\delta\mathbf{v})$ term from the continuity equation and the $(\mathbf{v} \cdot \nabla)\mathbf{v}$ term from the momentum equation have been ignored.

In a material coordinate approach, an approximation is formed once a particular model for the $\Psi(\mathbf{q}, t)$ mapping is made. Naturally, agreement with (3.29) must be insured for small displacements. Once we have adopted a form for Ψ, it follows that we have adopted an equation of motion for each particle. The peculiar velocity, as a function of $\mathbf{q}$ and t, is given by

$$\mathbf{v} = a(t)\frac{\partial}{\partial t}\Psi(\mathbf{q}, t) \tag{3.33}$$

and the peculiar acceleration is

$$\mathbf{g} = \frac{\dot{a}}{a}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial t} = 2\dot{a}\frac{\partial \Psi}{\partial t} + a\frac{\partial^2 \Psi}{\partial t^2} \qquad (3.34)$$

The equation of continuity (3.11) will be satisfied exactly. The distributions of density and velocity in space are calculable once $\Psi(\mathbf{q}, t)$ is specified. That is, the mapping $\mathbf{x}(\mathbf{q}, t)$ will describe the entire picture of the motion. The Poisson equation relating $\nabla \cdot \mathbf{g}$ and δ will not be satisfied (hence the phrase "approximation"). However, for small displacements the required agreement with linear perturbation theory will insure that the Poisson equation will be satisfied to first order. Will an extrapolation beyond the linear regime work? The arguments presented above do not necessarily indicate that it would. On the other hand, that Zel'dovich was able to find a successful model should not come as a complete surprise.

Using the Zel'dovich mapping

$$\mathbf{x}(\mathbf{x}, t) = \mathbf{q} + b(t)\,\mathbf{u}(\mathbf{q})\,, \qquad (3.35)$$

the peculiar velocity as a function of $\mathbf{q}$ and t is

$$\mathbf{v} = a\dot{b}\,\mathbf{u} \qquad (3.36)$$

and for small displacements we require this to agree with the velocity of linear perturbation theory. Comparing the time dependence in equations (3.30) and (3.36), we see that $b(t)$ is the growing mode (e.g. $b \sim t^{2/3}$ for $\Omega = 1$) solution of

$$\ddot{b} + 2\frac{\dot{a}}{a}\dot{b} - 4\pi G\bar{\rho} b = 0 \qquad (3.37)$$

Comparing equations (3.30) and (3.36) for small displacements, we may write

$$\mathbf{u}(\mathbf{q}) = -\nabla_q \Phi(\mathbf{q})\,, \qquad (3.38)$$

where Φ is independent of time and is proportional to the linear peculiar gravitational potential φ. Specifically, we have (note $\Delta \sim b$ and $\varphi \sim b/a$),

$$\Phi(\mathbf{q}) = \frac{1}{b}\Delta = \frac{1}{4\pi G\rho a^2 b}\varphi = \frac{2}{3H^2\Omega a^2 b}\varphi \tag{3.39}$$

If we factor out the time dependence in φ by writing $\varphi = b/a\,\varphi_0$ so that $\varphi_0(\mathbf{q}) = \varphi(\mathbf{x}, t_1)$ is the initial potential, then the relation of the Lagrangian mapping, with $\mathbf{u} = -\nabla_{\mathbf{q}}\Phi$, to the gravitational potential is particularly simple,

$$\Phi(\mathbf{q}) = \frac{1}{4\pi G\bar{\rho}a^3}\varphi_0 = \frac{2}{3H^2\Omega a^3}\varphi_0 \tag{3.40}$$

Since the velocity is a gradient in $\mathbf{q}$ space, it is curl free in material coordinates. It is more important from a physical standpoint to note that the spatial velocity field (in $\mathbf{x}$ coordinates) is irrotational, which we show below.

The deformation of a material element is described by the tensor $\partial\mathbf{x}/\partial\mathbf{q}$ which by equation (3.35) is

$$J_{ij} = \frac{\partial x^i}{\partial q^j} = \delta_{ij} + b\frac{\partial u^i}{\partial q^j} \tag{3.41}$$

and is symmetric, since $\mathbf{u}$ is a gradient. The Jacobian determinant J is of fundamental importance since, by equation (3.11), it directly determines the density in material coordinates. We may calculate it by expanding the determinant as an antisymmetric product

$$J = \frac{1}{3!}\epsilon_{ijk}\epsilon_{lmn}\left(\delta_{il} + b\frac{\partial u^i}{\partial q^l}\right)\left(\delta_{jm} + b\frac{\partial u^j}{\partial q^m}\right)\left(\delta_{kn} + b\frac{\partial u^k}{\partial q^n}\right), \tag{3.42}$$

which we can write in powers of b after using the following ϵ_{ijk} properties

$$\epsilon_{ijk}\epsilon_{ijk} = 3! \qquad \epsilon_{ijk}\epsilon_{ljk} = 2\delta_{il} \quad \text{and} \quad \epsilon_{ijk}\epsilon_{lmk} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} \tag{3.43}$$

Thus, the determinant as a function of $\mathbf{q}$ and t is

$$J = 1 + b\frac{\partial u^i}{\partial q^i} + b^2\frac{1}{2}\left(\frac{\partial u^i}{\partial q^i}\frac{\partial u^j}{\partial q^j} - \frac{\partial u^i}{\partial q^j}\frac{\partial u^j}{\partial q^i}\right) + b^3\left\|\frac{\partial \mathbf{u}}{\partial \mathbf{q}}\right\| \tag{3.44}$$

We can explicitly check that the density obtained from equation (3.11) agrees to first order with linear perturbation theory results. The material coordinate density contrast in the Zel'dovich Approximation is

$$\delta = J^{-1} - 1 = -b(t)\,\frac{\partial u^i}{\partial q^i} + \mathcal{O}(b^2) \tag{3.45}$$

This agrees with equation (3.29) since δ separates into a time dependent factor $b(t)$ and coordinate factor giving the primordial density field

$$\delta_0(\mathbf{q}) = -\nabla_{\mathbf{q}} \cdot \mathbf{u} \tag{3.46}$$

The above equation is in material coordinates but it conveys the same information to lowest order as the linearized spatial coordinate density velocity relation (from continuity), with $\delta = b(t)\,\delta_0(\mathbf{x})$,

$$\begin{aligned} \frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot \mathbf{v} &= 0 \\ \delta = -\frac{b}{\dot{b}a}\nabla \cdot \mathbf{v} &= -\frac{1}{faH}\nabla \cdot \mathbf{v} \end{aligned} \tag{3.47}$$

We have shown that the mapping (3.35) is consistent with the velocity and density of linear perturbation theory.

To quote the master Zel'dovich himself "the approximation proposed in this article consists in the extrapolation of formula [(3.28)] into the region where the perturbations in density [δ] are not small" (Zel'dovich 1970, p. 84). This paper is reprinted in his Selected Works, Volume II Particles, Nuclei, and the Universe (Zel'dovich 1993) which spans his pre-World War II Uranium fission work through the cosmology papers written in his last few decades.

The Zel'dovich approximation is a kinematic model which retains important features of the dynamics. The asymmetric nature of gravitational instability—pancake formation—is predicted in a cosmological context. The Eulerian velocity field remains curl free. The peaks in the linear density field act as attractors. But the nature of the material coordinate deformations mimics the tidal interaction of neighboring masses and thus affects the shape (primarily flattened) and connectivity (cellular structure) of the structures formed. The nature of perturbation theory in material coordinates allows for a simple scheme like the Zel'dovich mapping to produce a second order Eulerian perturbation

$$\delta^{(2)} = \frac{1}{2}(\delta^{(1)})^2 + \delta_{,i}^{(1)}\Delta_{i}^{(1)} + \frac{1}{2}\Delta_{ij}^{(1)}\Delta_{ij}^{(1)}, \tag{3.48}$$

which we show below. The nonlocal effects are nearly those of the exact theory—compare (3.48) with the exact second order expression (2.166).

3.2 Pancakes and Other Features

The first and most important consequence of the Zel'dovich mapping is the formation of pancakes—relatively thin sheet-like structures of high density. Due to the continuity and smoothness of the mapping, the pancakes (also known as caustics) tend to be somewhat connected and thus form a cellular structure. The approximation's pancaking property may be thought of as a virtue because of the accumulating observational evidence that the large-scale matter distribution contains bubbles, walls, and voids. The Poisson equation is not grossly violated, even when the approximation has been extended to the point of pancake formation, when δ is significantly greater than 1 in the dense regions. The proposed $\mathbf{x}(\mathbf{q},t)$ mapping does a good job of representing gravitational instability in an expanding universe. In fact, the approximate solution shares several properties, both quantitative and qualitative, with the exact solution. These points become clearer as we present some of the details below.

Let us define the material coordinate deformation tensor as

$$d_{ij} = \frac{\partial w^i}{\partial q^j} = -\frac{\partial^2 \Phi}{\partial q^i \partial q^j}, \tag{3.49}$$

which is symmetric by construction. Consider a cubical material element at initial location $\mathbf{q}$ (i.e. consider a group of particles near particle $\mathbf{q}$), occupying volume d^3q with the cube specified by the three perpendicular lengths dq^i. In general, the material element undergoes deformation, and the cube bounded by dq^i transforms into a parallelepiped bounded by

$$dx^i = \frac{\partial x^i}{\partial q^j} dq^j \tag{3.50}$$

occupying volume $\|\partial\mathbf{x}/\partial\mathbf{q}\| d^3q$. At each point $\mathbf{q}$, the deformation tensor may be diagonalized, producing three real eigenvalues corresponding to three mutually perpendicular eigenvectors. We denote the three eigenvalues of d_{ij} by $-\lambda_i$. The negative sign is added by convention, following Shandarin and Zel'dovich (1989), so the λ_i are eigenvalues of $\Phi_{ij} = \partial\Phi/\partial q^i\partial q^j$, which describes the curvature of the initial linear gravitational potential. Without loss of generality we order the eigenvalues by

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \tag{3.51}$$

In coordinates which diagonalize d_{ij} at $\mathbf{q}$, the q^i define the principle axes, and the deformation is described by the matrix

$$J_{ij} = \frac{\partial\mathbf{x}}{\partial\mathbf{q}} = \begin{pmatrix} 1 - b(t)\lambda_1(\mathbf{q}) & 0 & 0 \\ 0 & 1 - b(t)\lambda_2(\mathbf{q}) & 0 \\ 0 & 0 & 1 - b(t)\lambda_3(\mathbf{q}) \end{pmatrix} \tag{3.52}$$

If the initial cube is oriented so that its sides dq^i coincide with the new coordinate axes—the eigenvectors of d_{ij}—then the deformation is particularly simple to describe: the cube suffers a pure compression or expansion in each of the q^i directions. Thus the cube transforms into a rectangular parallelepiped.

The tensor Φ_{ij} possesses three characteristic invariants

$$\begin{aligned} I_1 &= \lambda_1 + \lambda_2 + \lambda_3 \\ I_2 &= \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1 \\ I_3 &= \lambda_1\lambda_2\lambda_3 \end{aligned} \tag{3.53}$$

These invariants are indeed coordinate independent, which is shown by the fact that we can rewrite them in terms of tensor contractions

$$
\begin{aligned}
I_1 &= \Phi_{,ii} \\
I_2 &= \frac{1}{2}(\Phi_{,ii}\Phi_{,jj} - \Phi_{,ij}\Phi_{,ij}) = \frac{1}{2}I_1^2 - \frac{1}{2}\Phi_{,ij}\Phi_{,ij} \\
I_3 = \|\Phi_{,ij}\| &= \frac{1}{6}\Phi_{,ii}^3 - \frac{1}{2}\Phi_{,ii}\Phi_{,jk}\Phi_{,jk} + \frac{1}{3}\Phi_{,ij}\Phi_{,jk}\Phi_{,ki} \\
&= I_1 I_2 - \frac{1}{3}I_1^3 + \frac{1}{3}\Phi_{,ij}\Phi_{,jk}\Phi_{,ki}
\end{aligned}
\tag{3.54}
$$

The Jacobian is easily calculated from equation (3.52) to be

$$
\begin{aligned}
J &= (1 - b\lambda_1)(1 - b\lambda_2)(1 - b\lambda_3) \\
&= 1 - bI_1 + b^2 I_2 - b^3 I_3
\end{aligned}
\tag{3.55}
$$

Therefore the material coordinate density can be written in the particularly revealing form

$$
1 + \delta(\mathbf{q}, t) = \frac{1}{(1 - b\lambda_1)(1 - b\lambda_2)(1 - b\lambda_3)}
\tag{3.56}
$$

while trajectories do not cross. In the event of multistream flow, there are modifications (Shandarin & Zel'dovich 1989, Kofman, Bertschinger, Gelb, Nusser, & Dekel 1994), following from the fact that a sum over the various $\mathbf{q}$ values whose trajectories arrived at $\mathbf{x}$ would need to be included on the right hand side of equation (3.7).

Equation (3.56) is important, as it shows that collapse to infinite density is inevitable in regions where there is at least one positive eigenvalue λ_i. If, as is almost always the case, we have unequal eigenvalues, labeled $\lambda_1 > \lambda_2 > \lambda_3$, then if $\lambda_1 > 0$ the collapse is one dimensional so that deformation is primarily along the λ_1 eigenvector direction, giving a $\delta \to \infty$ caustic with the eigenvector as its normal. After the pancake collapse, there follows, if $\lambda_2 > 0$, compression within the sheet which tends to form filamentary structure. Finally, if λ_3 is also positive, collapse along filaments produces clusters. These features are qualitatively born out in complete N-body simulations of structure formation in an expanding universe (Kofman, Pogosyan, Shandarin, & Melott 1992). As an illustration we produce in Figure 3-1 a somewhat generic numerical realization of the Zel'dovich approximation in three dimensions. We have plotted the mass distribution in a thin slice.

A major problem of the approximation stems from the fact that it is ballistic and so the pancakes dissolve—the smaller scale gravitational interactions are effectively turned off and particles do not turn around and fall into the potential wells. The blessing given to the approximation, however, is directly related to addressing this problem. Gravitational instability within an expanding spacetime can be successfully analyzed with spatial filters, thus washing out the small scale problems of the Zel'dovich approximation (Kofman et al. 1992). The weaknesses of the approximation are thus somewhat obvious, and in comparing various approximation schemes of gravitational instability (some of which we mention below) with N-body results, the Zel'dovich approximation fares surprisingly well (Coles, Melott, & Shandarin 1993). The cross correlations between Zel'dovich and N-body simulated data are quite encouraging—provided that the initial fluctuation spectrum in the Zel'dovich simulations has been filtered. The surprise is that evolution of the smaller scale substructure does not skew the large-scale results in such a way that successful filtering breaks down on all but the largest (linear regime) scales. As a result, the formal title "Truncated Zel'dovich Approximation" is sometimes used in the literature to reflect the successful usage of filtering (the "truncation" refers to large k suppression of the $\langle|\delta_0(\mathbf{k})|^2\rangle$ spectrum.

The baryonic pancake and neutrino HDM models have faded. These models utilized the Zel'dovich approximation in a top-down sense to generate the large-scale cellular structure (cf. Doroshkevich, Ryaben'kii, & Shandarin 1973; Shandarin, Doroshkevich, & Zel'dovich 1983). A short review which indicates the general philosophy of such models is contained in a collection of review articles by Zel'dovich (1992).

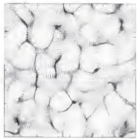


Figure 3-1 A two dimensional slice from a three dimensional numerical realization of the Zel'dovich approximation

A bottom-up hierarchical paradigm was developed by Peebles and others during this same period (cf. Peebles 1974a, Peebles 1980). The hierarchical models had closer ties with the ≈ 1–$10\,h^{-1}$ Mpc scale correlation data. The Press-Schechter (1974) algorithm for obtaining a bottom-up mass spectrum of collapsed objects based on the linear density field is a practical hierarchical example. With closer collaboration between the two schools of thought during the 1980's, coupled with the rapid improvements in computer simulations, there has been a merger of sorts. Qualitatively the bottom-up features of the hierarchical CDM models have survived. Utilizing filtering in Lagrangian mapping approximation schemes appears to give semi-quantitative insight into the formation of large coherent structures, which can be called super-pancakes and super-filaments (Kofman et al. 1992). Such structures appear in the simulation data as well as in redshift surveys.

We now turn our attention to some of the features of the Zel'dovich approximation. Once we have the Zel'dovich map, it is easy to check that it obeys the Eulerian $\nabla \times \mathbf{v} = 0$ condition. The curl-free condition (3.24) holds identically, which we observe by using the frame where J_{ij} is diagonal, but note that the condition becomes indeterminate (we divide by zero) once caustic formation occurs, i.e. we need the quantity $1 - b(t)\lambda_1(\mathbf{q})$ to be positive at the point $\mathbf{x}(\mathbf{q}, t)$ where the curl is taken.

The Zel'dovich approximate solution is actually the exact solution in one dimension, prior to multistreaming. That is, in the artificial case in which perturbations depend on one coordinate $\delta(x, y, z) \equiv \delta(x)$, corresponding to a set-up for pancake formation parallel to the yz-plane, the exact dynamics in an expanding background gives the Zel'dovich mapping. This is easy to check. The non-dimensional mapping gives

$$x = q + b\,u$$
$$\delta = J^{-1} - 1 = \frac{-b\,\partial u/\partial q}{1 + b\,\partial u/\partial q} \tag{3.57}$$

The peculiar acceleration in one, two or three dimensions is

$$\begin{aligned} \boldsymbol{g} &= \frac{d\mathbf{v}}{dt} + \frac{\dot a}{a}\mathbf{v} \\ &= (2\dot a\dot b + a\ddot b)\,\mathbf{u} = 4\pi G\bar\rho a b\,\mathbf{u} \end{aligned} \tag{3.58}$$

In one dimension, we may simply compute the divergence of $\boldsymbol{g}$ to find

$$\nabla\cdot\boldsymbol{g} = \frac{\partial g}{\partial q}\frac{\partial q}{\partial x} = -4\pi G\bar\rho a\delta \tag{3.59}$$

with δ given by (3.57), so we see that the Poisson equation is satisfied exactly.

In two or three dimensions, the Poisson equation is only met to first order. We may keep track of the fractional error in the Poisson equation in the following manner (Shandarin & Zel'dovich 1989). Let us define δ_* to be the dynamical density calculated from the divergence of $\boldsymbol{g}$ in the Zel'dovich approximation,

$$\delta_* = -\frac{1}{4\pi G\bar\rho a}\nabla\cdot\boldsymbol{g}, \tag{3.60}$$

and we define the fractional error as

$$\text{error} = \frac{(1+\delta_*)-(1+\delta)}{1+\delta} \tag{3.61}$$

where δ is $J^{-1}-1$ from the material continuity equation. For the divergence of $\boldsymbol{g}$ we write, using (3.58)

$$\begin{aligned} -\frac{1}{4\pi G\bar\rho a}\nabla\cdot\boldsymbol{g} &= -b\frac{\partial u^j}{\partial q^k}\frac{\partial q^k}{\partial x^j} \\ &= b\left[\frac{\lambda_1}{1-b\lambda_1} + \frac{\lambda_2}{1-b\lambda_2} + \frac{\lambda_3}{1-b\lambda_3}\right], \end{aligned} \tag{3.62}$$

where we have diagonalized the deformation tensor. Using expression (3.56) for δ we obtain

$$\text{error} = -b^2 I_2 + 2b^3 I_3, \tag{3.63}$$

which grows with time but which remains finite even at caustic formation, since δ and δ_* diverged in the same manner. Equation (3.63) shows that the Poisson equation is satisfied to first order. Also, in the one-dimensional case, both λ_2 and λ_3 vanish, so that $I_2 = I_3 = 0$ and there is no error, as we showed above.

When $\delta \rightarrow \infty$ in the approximation, we still have a finite mass per unit area in the sheet of collapsed matter, since the perturbation deforms one-dimensionally and the gravitational potential is finite and smooth, and the error remains finite even at collapse. The finite nature of the singularity at a caustic may seen from generic considerations of smooth Lagrangian maps. Consider a pancake forming at $\mathbf{x}_0 = \mathbf{x}(\mathbf{q}_0, t_0)$. We diagonalize the deformation tensor J_{ij} at $\mathbf{q}_0$ as in equation (3.52). A sheet is forming at time given by $1 - b(t_0)\lambda_1 = 0$. We reduce to the one dimensional problem for simplicity, with no essential loss of generality (Shandarin & Zel'dovich 1989), and we let x be the λ_1 eigenvector direction. For material coordinates $q = q_0 + \Delta q$ and Eulerian coordinates $x = x_0 + \Delta x$ near the sheet, we have

$$\begin{aligned} J(q) &= J(q_0) + \frac{\partial J}{\partial q}\Delta q + \frac{1}{2}\frac{\partial^2 J}{\partial q^2}\Delta q^2 + \cdots \\ x &= x(q_0) + \frac{\partial x}{\partial q}\Delta q + \frac{1}{2}\frac{\partial^2 x}{\partial q^2}\Delta q^2 + \frac{1}{6}\frac{\partial^3 x}{\partial q^3}\Delta q^3 + \cdots \end{aligned} \tag{3.64}$$

If λ_1 is a relative maximum, then in the single-stream flow at t_0 near x_0, we have $J(q) \sim \Delta q^2$ and $\Delta x \sim \Delta q^3$, which gives

$$1 + \delta(\mathbf{x}) \sim |\mathbf{x} - \mathbf{x}_0|^{-2/3} \tag{3.65}$$

as the density profile near the forming caustic. Slightly after t_0, we have triple-stream flow. The $u(q)$ function is smooth, and it remains smooth with respect to the Eulerian coordinates x, but $\partial u/\partial x$ goes to $-\infty$ at x_0, immediately after which the pancake at x_0 corresponds to two singularities, near x_0, between which there is triple-stream flow. In that triple-stream case (at $t_0 + \Delta t$), and in the more generic case where δ is blowing up but we do not have a relative maximum of λ_1 at q_0, there is also a manageable singularity. We have $J(q) \sim \Delta q$ and $\Delta x \sim \Delta q^2$, which gives

$$1 + \delta(\mathbf{x}) \sim |\mathbf{x} - \mathbf{x}_s|^{-1/2} \tag{3.66}$$

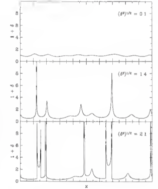


Figure 3-2 The growth of density fluctuations with time: the formation of pancakes in a one dimensional simulation. The linear evolution $\langle\delta^2\rangle \sim b^2$ gives $\langle\delta^2\rangle^{1/2}$ values of 0.1, 1 and 1.5; the actual values from the simulation (noted on the diagram) are larger due to nonlinear evolution.

as the generic density profile near a singularity at $\mathbf{x}_s$. These cusps may be integrated over to obtain a finite mass per unit area in a pancake. The general description of the singularities of the Zel'dovich mapping in one and two dimensions was analyzed by Arnold, Shandarin and Zel'dovich (1982) and placed within the context of catastrophe theory by Arnold (1992).

We have performed one-dimensional numerical simulations which clarify the above remarks. In Figure 3-2, we have evolved some rather generic initial conditions to display caustic formation. The initial $\langle|\delta_0(k)|^2\rangle$ spectrum was simply flat, with a cutoff at large k. The top density profile is in the linear regime, with $\langle\delta^2\rangle^{1/2} = 0.1$ and the bottom profiles have $\langle\delta^2\rangle \gtrsim 1$, as shown. The single- and triple-stream regions near caustics are evident in this picture. In Figures 3-3 and 3-4, we have zoomed in on a pancake from another generic realization of the approximation. The Eulerian velocity field $v/a\dot{b}$ is shown to be single-valued in Figure 3-3, with $\partial v/\partial x \to -\infty$ at the cusp. In Figure 3-4, the velocity is triple-valued in the region of the thickening pancake. Outside of multistream regions, the approximation still corresponds to the exact solution of one-dimensional instability in an expanding universe.

We performed two-dimensional simulations which show networks of pancakes evolving from generic initial conditions (again, flat $\langle|\delta_0(\mathbf{k})|^2\rangle$ with high k cutoff). In Figure 3-5 we show a typical example, evolved until $\langle\delta^2\rangle \approx 1$. We show a close-up view of the thickening of several pancakes in Figure 3-6.

The statistics and form of the initial gravitational potential clearly provide the seeds for later structure formation. In view of this, Doroshkevich (1970) wrote a seminal paper concerned with the spatial structure of an initial random field. This was a precursor to the famous BBKS (Bardeen, Bond, Kaiser & Szalay 1986) paper on Gaussian random fields in cosmology. As the Zel'dovich approximation clearly

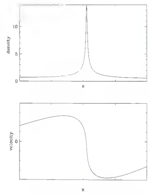


Figure 3-3 Generic kinematics of the single stream Eulerian velocity field v/ab perpendicular to a pancake from a one-dimensional simulation

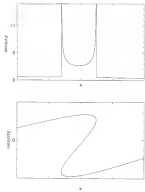


Figure 3-4 The pancake thickens. Generic kinematics of the triple stream Eulerian velocity field v/ab perpendicular to a pancake from a one-dimensional simulation

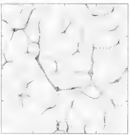


Figure 3-5 A typical two-dimensional simulation. A somewhat connected network of pancakes is formed

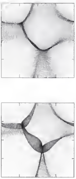


Figure 3-6 A closer view of smooth collapse into pancakes. The top picture, which has single-stream flow at the caustics, was evolved until $\langle\delta^2\rangle^{1/2} \approx 1$. The bottom picture is further evolved, $\langle\delta^2\rangle^{1/2} \approx 1.5$, and we see triple-stream flow in the pancakes.

shows it is not simply the peaks in the linear density field which survive as later structure. From the expansion of the determinant J we see that to first order

$$\delta(\mathbf{x}, t) = b(t)\,(\lambda_1 + \lambda_2 + \lambda_3) = b(t)\,\nabla^2\Phi \tag{3.67}$$

and therefore we have placed deformation information into the equation from the beginning. The clearest counter-example to Eulerian linear theory is the case where $\lambda_1 > 0$ yet $\sum \lambda_i < 0$. There is pancake formation in a region which would become increasingly underdense according to linear evolution.

For an initial $\delta_0(\mathbf{q})$ field which is Gaussian (so the initial potential, velocity and density fields are all Gaussian), with $\sigma_0^2 = \langle \delta_0^2 \rangle$, Doroshkevich found the joint probability distribution for the set

$$\lambda_1 > \lambda_2 > \lambda_3 \tag{3.68}$$

of eigenvalues of the deformation tensor $\Phi_{,ij}$

$$\mathcal{P}(\lambda_1, \lambda_2, \lambda_3) = \frac{5^{5/2}27}{8\pi\sigma_0^6}(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)\exp\left[-\frac{3I_1^2 - 15I_2/2}{\sigma_0^2}\right] \tag{3.69}$$

This implies that the probability that one value is positive while the other two are negative (which is equal to the probability that two are positive and one is negative) is $\approx 42\%$. The probability that all three are positive (which is the same as that for all three negative) is $\approx 8\%$. In a purposefully naive approach, we might ignore the ordering (3.68) and that the λ_i are dependent (since they come from the $\Phi_{,ij}$ eigenvalue problem) random variables. Then we would estimate that the probability that all the λ_i are negative is simply $(1/2)^3 = 12.5\%$. That the answer (8%) is less is due to the property of the repulsion of eigenvalues (Zel'dovich 1983).

As explained by Shandarin and Zel'dovich (1989), the high density regions in the evolved, non Gaussian density field percolate, despite the fact that the dense regions themselves comprise but a small volume of space. Eventual sheets, filaments and clusters correspond to $\lambda_1 > 0$ regions comprising $\approx 92\%$ of q-space by the above calculation by Doroshkevich The percolation of the deformation properties in q-space go over into x space, since the $\mathbf{x}(\mathbf{q}\ t)$ mapping is smooth Consider the smoothness of the initial conditions and the agreement of the expansion time scale with the instability time scale These two points along with the asymmetric nature of gravitational instability are combined to the Zel dovich approximation to yield cosmological pan cakes. The continuity/percolation argument above then gives the cellular structure further details of such an argument are provided by Sahni Sathyaprakash, and Shandarin (1994) Hierarchical structure formation allows for small scales to be nonlinear while large scales can usefully be analyzed with linear theory It is interesting that approximations to gravitational instability with suitable filtering are able to bridge the gap somewhat, between the the linear and nonlinear regimes

3.3 Perturbation Theory

3.3.1 Perturbation Theory in the Zel dovich Approximation

The Zel'dovich approximation has its own perturbation theory similar to structure to the full gravitational equations Within the confines of this approximation there is no need to provide a perturbation series in $\mathbf{v}$, the peculiar velocity because naturally, the relation $\mathbf{v} = a\dot{\mathbf{u}}$ for the material coordinate velocity field holds throughout The Eulerian and Lagrangian density fields are expanded based on the exact mass conservation relation $1 + \delta = J^{-1}$ and the $\mathbf{x}(\mathbf{q}\ t)$ mapping and we may obtain the Eulerian field

$$\delta(\mathbf{x}, t) = \delta^{(1)} + \delta^{(2)} + \tag{3 70}$$

More generally the material displacement mapping $\Psi(\mathbf{q}\ t)$ can be analyzed perturbatively from the start

$$\Psi(\mathbf{q}\ t) = \Psi^{(1)} + \Psi^{(2)} + \tag{3 71}$$

going beyond the Zel'dovich approximation (the linear term $\Psi^{(1)} = b\mathbf{u}$) if needed. In this approach, called Lagrangian perturbation theory, the material coordinate trajectories are the primary object, rather than the Eulerian density contrast field. Relating the Lagrangian and Eulerian approaches order by order shows agreement, but the Lagrangian method may always be pushed considerably further, since large δ values are not necessarily forbidden, as first shown by Zel'dovich. Perturbation theory in the sense of equation (3.71) is reviewed by Bouchet, Colombi, Hivon, and Juszkiewicz (1995).

The nature and validity of the Zel'dovich approximation was clarified by Grinstein and Wise (1987), who initiated the technique whereby approximation schemes are tested in the quasi-linear regime according to their perturbation theory. Further progress along these lines, clarifying the Lagrangian method in general, and its relation to Eulerian perturbation theory, was presented by Bernardeau, Singh, Banerjee, and Chitre (1994) and Munshi, Sahni, and Starobinsky (1994). The initial developments of the cosmological Lagrangian perturbation theory were made by Zel'dovich (1970), Moutarde, Alimi, Bouchet, Pellat, and Ramani (1991), Buchert (1992), and Bouchet et al. (1992).

We now turn to the perturbations induced by the Zel'dovich approximation. Going to second order in Zel'dovich approximation perturbation theory is straightforward, as it is based on the expansion of the deformation Jacobian $J = \|\partial\mathbf{x}/\partial\mathbf{q}\|$ in equation (3.44). The density expansion in material coordinates is

$$\begin{aligned}\delta_m = J^{-1} - 1 &= -b\frac{\partial u^i}{\partial q^i} + b^2\frac{1}{2}\left[\left(\frac{\partial u^i}{\partial q^i}\right)^2 + \frac{\partial u^i}{\partial q^j}\frac{\partial u^j}{\partial q^i}\right] + \mathcal{O}(b^3) \\ &= \delta_m^{(1)} + \delta_m^{(2)} + \mathcal{O}(b^3),\end{aligned} \qquad (3.72)$$

where we use a subscript m for the material coordinate perturbations, since $\delta_m^{(i)} \neq \delta^{(i)}$ in general. In single-stream mode, we may connect $\mathbf{x}$ and $\mathbf{q}$ for some function $f(\mathbf{q})$

with the Taylor series

$$f(\mathbf{x}) = f(\mathbf{q}) + b u^j \frac{\partial}{\partial q^j} f(\mathbf{q}) + \mathcal{O}(b^2) \tag{3.73}$$

so we have, for example

$$\frac{\partial u^i}{\partial q^j} = \frac{\partial u^i}{\partial x^j} - b u^k \frac{\partial}{\partial x^k}\left(\frac{\partial u^i}{\partial x^j}\right) + \mathcal{O}(b^2) \tag{3.74}$$

Thus, we may get the first two terms of the Eulerian density field

$$\delta^{(1)} = -b \frac{\partial u^j}{\partial x^j} \tag{3.75}$$

$$\delta^{(2)} = b^2 \left[\frac{1}{2}\left(\frac{\partial u^j}{\partial x^j}\right)^2 + u^k \frac{\partial}{\partial x^k}\left(\frac{\partial u^j}{\partial x^j}\right) + \frac{1}{2}\frac{\partial u^k}{\partial x^j}\frac{\partial u^j}{\partial x^k} \right] \tag{3.76}$$

We may relate this to our previous Eulerian perturbation theory notation, whereby equation (3.75) simply reproduces $\delta^{(1)} = b\,\delta_0(\mathbf{x})$, with a linear peculiar velocity field of $\mathbf{v}/ab = -\nabla\Phi = -\nabla\Delta^{(1)}/b$. Writing quantities in (3.76) using only the linear density and potential fields, we have

$$\delta^{(2)} = \frac{1}{2}(\delta^{(1)})^2 + \delta^{(1)}_{,i}\Delta^{(1)}_{,i} + \frac{1}{2}\Delta^{(1)}_{,ij}\Delta^{(1)}_{,ij} \tag{3.77}$$

as we mentioned previously.

With Zel'dovich approximation perturbation theory, it is possible to go further than this, as first shown by Grinstein and Wise (1987). The Zel'dovich approximation is inherently set in within the Lagrangian formulation. To specify the approximation in the Eulerian picture, we generate the full perturbation series with a recursion for the sequence of kernels $G_n^{(s)}(\mathbf{k}_1, \ldots, \mathbf{k}_n)$ used in the k-space perturbation expansion, which is the Fourier transform of (3.70). In fact, for the Zel'dovich mapping, the general formula turns out to be nice and concise, as we now show, since it is important to have a worked out example of the Eulerian/Lagrangian relation.

The Lagrangian to Eulerian coordinate mapping is

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + b(t)\mathbf{u}(\mathbf{q}), \tag{3.78}$$

and before orbit crossing we have an inverse $\mathbf{q}(\mathbf{x}, t)$ map as well as a single-valued Eulerian velocity field $\mathbf{u}_s(\mathbf{x}, t) = \mathbf{v}(\mathbf{x}, t)/a\dot{b}$ where the subscript s specifies that this is the spatial as opposed to material coordinate velocity. The Eulerian distribution function is well defined such that

$$f(\mathbf{x}, \mathbf{u}_s, t)\, d^3x\, d^3u_s \tag{3.79}$$

is the number of particles (of equal mass) in the phase space volume $d^3x\, d^3u_s$ at time t. The initially uniform spatial distribution (number density $\bar{n}$) evolves with time according to

$$f(\mathbf{x}, \mathbf{u}_s, t) = \bar{n}\delta_D(\mathbf{u}_s - \mathbf{u}(\mathbf{q})) \tag{3.80}$$

where $\mathbf{q} = \mathbf{x} - b\mathbf{u}_s$. Since the mapping is smooth and single-valued, we expand in a Taylor series,

$$\begin{aligned} f(\mathbf{x}, \mathbf{u}_s, t) &= \bar{n}\left[\delta_D(\mathbf{u}_s) - \mathbf{u}\cdot\nabla_{\mathbf{u}_s}\delta_D(\mathbf{u}_s) + \frac{1}{2}u^i u^j \frac{\partial}{\partial u_s^i}\frac{\partial}{\partial u_s^j}\delta_D(\mathbf{u}_s) - \cdots\right] \\ &= \bar{n}\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}u^{j_1}\cdots u^{j_n}\frac{\partial}{\partial u_s^{j_1}}\cdots\frac{\partial}{\partial u_s^{j_n}}\delta_D(\mathbf{u}_s) \end{aligned} \tag{3.81}$$

The density is given by

$$1 + \delta(\mathbf{x}, t) = \frac{1}{\bar{n}}\int f(\mathbf{x}, \mathbf{u}_s, t)\, d^3u_s \tag{3.82}$$

and if we use the fact that

$$\int f(x)\frac{d^n}{dx^n}\delta_D(x)\, dx = (-1)^n\frac{d^n}{dx^n}f(0)\,,$$

we have for the Eulerian density

$$\begin{aligned} \delta(\mathbf{x}, t) &= \nabla_{\mathbf{u}_s}\cdot\mathbf{u}(\mathbf{x} - b\mathbf{u}_s)\Big|_{\mathbf{u}_s=0} + \frac{1}{2}\frac{\partial}{\partial u_s^i}\frac{\partial}{\partial u_s^j}u^i(\mathbf{x} - b\mathbf{u}_s)u^j(\mathbf{x} - b\mathbf{u}_s)\Big|_{\mathbf{u}_s=0} + \cdots \\ &= -b\nabla_{\mathbf{x}}\cdot\mathbf{u}(\mathbf{x}) + \frac{1}{2}b^2\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j}u^i(\mathbf{x})u^j(\mathbf{x}) - \cdots \end{aligned} \tag{3.83}$$

which give us our final Eulerian result. Thus the Zel'dovich approximation x-space perturbation series can be conveniently written as

$$\delta(\mathbf{x},t) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} b^n \frac{\partial}{\partial x^{i_1}} \cdots \frac{\partial}{\partial x^{i_n}} \left[u^{i_1}(\mathbf{x}) \cdots u^{i_n}(\mathbf{x})\right] \tag{3.84}$$

Note that the first two terms reproduce (3.75) and (3.76). Consider the Fourier transform of (3.84)

$$\delta(\mathbf{k},t) = \int \frac{d^3x}{V} e^{-i\mathbf{k}\cdot\mathbf{x}} \delta(\mathbf{x},t) \tag{3.85}$$

With the velocity field in (3.84) written as its Fourier sum

$$\mathbf{u} = \int \frac{V d^3k'}{(2\pi)^3} e^{i\mathbf{k}'\cdot\mathbf{x}}\, i\frac{\mathbf{k}'}{k'^2} \delta_0(\mathbf{k}') \tag{3.86}$$

we see that the $\delta^{(n)}$ term in (3.85) will contain n modes $\delta^{(1)}(\mathbf{k}_1) \cdots \delta^{(1)}(\mathbf{k}_n)$, and the $\int d^3x$ integral in (3.85) gives the usual momentum conserving delta function at each order n,

$$[\delta_D] = \left[\frac{(2\pi)^3}{V} \delta_D(\mathbf{k} - \textstyle\sum \mathbf{k}_i)\right] \tag{3.87}$$

The derivatives in (3.84) pull down $i\mathbf{k}$ factors, and the k-space formalism works out compactly. The final result for the Zel'dovich approximation is therefore

$$\begin{aligned} \delta(\mathbf{k},t) &= \sum_{n=1}^{\infty} b^n \int \frac{V d^3k_1}{(2\pi)^3} \cdots \frac{V d^3k_n}{(2\pi)^3} [\delta_D] G_n(\mathbf{k}_1, \ldots, \mathbf{k}_n) \delta_0(\mathbf{k}_1) \cdots \delta_0(\mathbf{k}_n) \\ G_n(\mathbf{k}_1, \ldots, \mathbf{k}_n) &= \frac{1}{n!} \frac{(\mathbf{k}\cdot\mathbf{k}_1)}{k_1^2} \frac{(\mathbf{k}\cdot\mathbf{k}_2)}{k_2^2} \cdots \frac{(\mathbf{k}\cdot\mathbf{k}_n)}{k_n^2} \end{aligned} \tag{3.88}$$

where we set $\mathbf{k} = \sum \mathbf{k}_i$. Note that the kernel $G_n(\mathbf{k}_1, \ldots, \mathbf{k}_n)$ is already symmetrized. The first kernel in the sequence is $G_1 = 1$, which gives $\delta^{(1)} = b\delta_0$, as it should, while the second kernel is

$$G_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{1}{2} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{1}{2} \left(\frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2}\right)^2$$

which agrees with equation (3.77) and is similar to the exact gravitational kernel

3.3.2 Other Approximations

While the Zel'dovich approximation is used in this work, there are of course other variations to the quest to approximate gravitational instability beyond the linear regime. Many of the various methods are reviewed by Sahni and Coles (1995).

The Zel'dovich mapping plus glueing will at least qualitatively address the shell crossing problems of the approximation (Zel'dovich 1983). Gurbatov, Saichev and Shandarin (1989) formalized such a method, creating the elegant Adhesion model. With the addition of a small viscosity term to the Euler equation, pancake structures are maintained due to a thin viscous boundary layer. The pancake → filament → clump evolutionary sequence (when all the Φ_{ij} eigenvalues are positive) is maintained and structures such as voids evolve in an environment leading to super-pancakes and super-filaments (Kofman et al. 1992). The method is beautifully described by Sahni et al. (1994) and Sahni and Coles (1995).

If we use the formalism of Lagrangian perturbation from the beginning, using the perturbation expansion (3.71) for the $\mathbf{x}(\mathbf{q}, t)$ mapping, we obtain what are sometimes called the post-Zel'dovich approximation (second order) and the post-post-Zel'dovich approximation (third order). The material coordinate equations (3.24) and (3.25) need to be solved order by order in an iterative manner. To first order the solution for $\Psi^{(1)}$ is separable, leading to the Zel'dovich mapping. As shown by Bouchet et al. (1992) and Bouchet et al. (1995), the higher order solutions are separable if $\Omega = 1$ and nearly so when $\Omega \neq 1$, and the solution has remarkably little Ω dependence. The second order $\Psi^{(2)}$ solution for $\Omega = 1$ is determined by

$$\begin{aligned} \Psi^{(2)}_{i,i} &= -\frac{3}{14}\left[(\Psi^{(1)}_{i,i})^2 - \Psi^{(1)}_{i,j}\Psi^{(1)}_{j,i}\right] \\ \Psi^{(2)}_{j,i} &= \Psi^{(2)}_{i,j} \end{aligned} \qquad (3.83)$$

which gives the $\Psi^{(2)}$ part of the map since it is a gradient. Adding the (3.83) contribution to the $1/J$ expansion (3.72) for the density gives an Eulerian $\delta^{(2)}$ of the same form as equation (3.74), but the $\frac{1}{2}$, $\frac{1}{2}$ coefficients do get shifted to the exact values of $\frac{5}{7}$, $\frac{2}{7}$. To illustrate the near separability in the $\Omega \neq 1$ case, we note that the $\frac{5}{7}$, $\frac{2}{7}$ values go to $1 \pm \kappa$, where κ is well approximated by $\frac{3}{14}\Omega^{-2/63}$. The Zel'dovich approximation perturbation theory displays the same structure as the exact theory. Combine this with the filtered approximation's favorable cross correlations with N-body results (Coles et al. 1993), and we have motivation for further numerical experiments.

CHAPTER 4
SIMULATIONS

4.1 Numerical Realizations

4.1.1 The Grid

The Lagrangian mappings of material coordinate perturbation theory are particularly suited for numerical investigations using computers. The peculiar gravitational potential is modeled as a Gaussian random field initially, while fluctuations are small. This implies that the initial density fluctuation and peculiar velocity fields are also Gaussian. As we illustrated in Chapters 2 and 3, the potential φ, density contrast δ, and the velocity **v** are related by the equations of motion. By specifying the δ field as a simulated, pseudo-random Gaussian variable with specified 2 point spatial correlations, we have the first order fields which we may evolve in time according to perturbation theory. We go further by extrapolating the fluctuations into the nonlinear regime using the Zel'dovich approximation.

We discuss aspects of the numerical methods below. We show the evolution of simulations into the nonlinear regime, and show the agreement, during the early stages of evolution, with linear theory. To close the chapter, we present pictures which give a visual impression of the dependence of the final matter distribution on the initial correlations $P_0(k)$ and also to any biasing between the matter and the galaxy distributions.

First we clarify the definition of the linear fields and their relation in Fourier space. The density contrast is

$$\delta(\mathbf{x}, t) = b(t)\,\delta_0(\mathbf{x}) \tag{4.1}$$

The rescaled gravitational potential Φ is given by

$$\Phi(\mathbf{x}) = \frac{1}{4\pi G\bar{\rho}a^2 b}\varphi = \frac{1}{4\pi G\bar{\rho}a^3}\varphi_0 , \tag{4.2}$$

where φ_0 is the initial peculiar gravitational potential. The linear solutions are related by

$$\nabla^2\Phi = \delta_0 \tag{4.3}$$

$$\mathbf{u} = \mathbf{v}/a\dot{b} = -\nabla\Phi \tag{4.4}$$

The density fluctuations are specified in k-space, where the covariance function is diagonal. Since the fluctuations are independently distributed for different $\mathbf{k}$ values, it is straightforward to lay down a $\delta_0(\mathbf{k})$ field on a simulation grid in practice. The power spectrum $P_0(k)$, as we introduced it in Chapter 2, will specify the 2-point correlations and thus determines δ_0 completely since all higher order connected moments vanish. The x-space and k-space relation is

$$\delta(\mathbf{k}) = \int \frac{d^3x}{V} e^{-i\mathbf{k}\cdot\mathbf{x}}\,\delta(\mathbf{x}) \tag{4.5}$$

$$\delta(\mathbf{x}) = \int \frac{V d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}}\,\delta(\mathbf{k}) \tag{4.6}$$

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = \left[\frac{(2\pi)^3}{V}\delta_D(\mathbf{k}+\mathbf{k}')\right]P(k) , \tag{4.7}$$

where $P(k)$ is the power per mode, defined as

$$P(k) = \langle|\delta(\mathbf{k})|^2\rangle \tag{4.8}$$

The initial potential and velocities in k-space are

$$\begin{aligned} \Phi &= -\frac{\tilde{\delta}_0(\mathbf{k})}{k^2} \\ \mathbf{u} &= i\frac{\mathbf{k}}{k^2}\tilde{\delta}_0(\mathbf{k}) \end{aligned} \tag{4.9}$$

From the Fourier transform definitions, we see that the local variance of each linear field is proportional to an integral moment of the power spectrum

$$\int \frac{V d^3k}{(2\pi)^3} k^{2m} P(k), \tag{4.10}$$

with the $m = 0, -1$, and -2 moments corresponding to the variances of δ, $\mathbf{v}$, and φ, respectively. For the scaled initial fields δ_0, $\mathbf{u}$, and Φ, we express this as

$$\frac{d(\delta_0^2)}{d\ln k} = \frac{V 4\pi}{(2\pi)^3} k^3 P_0(k) \tag{4.11}$$

$$\frac{d(\mathbf{u}^2)}{d\ln k} = \frac{V 4\pi}{(2\pi)^3} k P_0(k) \tag{4.12}$$

$$\frac{d(\Phi^2)}{d\ln k} = \frac{V 4\pi}{(2\pi)^3} k^{-1} P_0(k) \tag{4.13}$$

If we wish to model the large scales by the Zel'dovich scale free spectrum, where potential fluctuations are the same on all scales, we choose $P_0 \sim k$, so that the Φ variance is scale free.

We implement a Lagrangian mapping on a discrete $n \times n \times n$ grid representing the comoving coordinates. A uniform grid represents initial material coordinates $\mathbf{q}$. After time evolution, the simulated data is analyzed with another grid of $\mathbf{x}$ coordinates. Fourier transforms are taken in the discrete sense, as noted in Chapter 2. To optimize the Discrete Fourier Transform (FFT) algorithm, the grid size n is a power of 2 such as 64 or 128. The fundamental cube is taken to be $L \times L \times L$ in comoving dimension. With grid spacing $\Delta x = L/n$, the grid coordinates are $\mathbf{x} = \mathbf{x}_{int}\Delta x$, where the $\mathbf{x}_{int}$ components are integers in the $0, 1, \ldots, n-1$ range. The comoving wavenumber $\mathbf{k}$ is associated with an $n \times n \times n$ lattice in k-space, with grid spacing $\Delta k = 2\pi/L$, i.e. $\mathbf{k} = \mathbf{k}_{int}\Delta k$ with integer $\mathbf{k}_{int}$ components in the range $0, 1, \ldots, n-1$, which maps to the range $-n/2+1, \ldots, n/2$ by the periodicity of the discrete series. The reality condition on the x space fields and the FFT symmetries imply that there are a total of n^3 independent real and imaginary components in the $k_z \geq 0$ half of k-space, so δ and $\tilde{\delta}$ contain the same quantity of information. The highest frequency (called the critical or Nyquist frequency), associated with an integer wavevector component of $n/2$, is $k_{nyq} = \pi/\Delta x$, and this corresponds to a wavelength of $2\Delta x$, i.e. critical sampling is two points per wave. At best, our frequency range in k-space will be limited by wavelengths of $2\pi/k = L$, the fundamental mode, to $2\pi/k = 2\Delta x$, the critical sampling length.

For purposes of the discrete numerical work, it proves convenient to work with the grid coordinates $\mathbf{x}_{int}$ and $\mathbf{k}_{int}$ directly, and we omit the int subscripts. The x-space k-space relation is

$$\tilde{\delta}(\mathbf{k}) = \frac{1}{n^3}\sum_{\mathbf{x}} \delta(\mathbf{x}) e^{-i2\pi\mathbf{x}\cdot\mathbf{k}/n} \tag{4.14}$$

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \tilde{\delta}(\mathbf{k}) e^{i2\pi\mathbf{x}\cdot\mathbf{k}/n} \tag{4.15}$$

$$\langle \tilde{\delta}(\mathbf{k})\tilde{\delta}(\mathbf{k}') \rangle = [\delta_D(\mathbf{k}+\mathbf{k}')] P(k) \tag{4.16}$$

where the [ ] delta function is either 0 or 1. We normalize the initial δ_0 field such that the $b(t)\delta_0(\mathbf{x})$ first order evolution corresponds to the measured $\sigma^2 = \langle \delta^2 \rangle$ and $P(k) = \langle |\tilde{\delta}(\mathbf{k})|^2 \rangle$ statistics in the linear regime. Such normalization doesn't require any adjustments; we need only insure that the linear perturbation theory relations among the fields hold true on the grid. We normalize the $\int d^3k k^{-2} P_0(k)$ integral out of convenience, since it is directly related to the Zel'dovich mapping. The velocity expansion on the grid, which satisfies $\nabla \cdot \mathbf{u} = -\delta_0$, is

$$\mathbf{u}(\mathbf{x}) = \sum_{\mathbf{k}} \frac{i\mathbf{k}}{\frac{2\pi}{n} k^2} \tilde{\delta}_0(\mathbf{k}) e^{i2\pi\mathbf{k}\cdot\mathbf{x}/n} \tag{4.17}$$

and we use this to compute the variance $\langle \mathbf{u}^2 \rangle$ which we then normalize to n^2

$$\langle \mathbf{u}^2 \rangle = \frac{n^2}{(2\pi)^2} \sum_{\mathbf{k}} \frac{|\delta_0(\mathbf{k})|^2}{k^2} \equiv n^2 \tag{4.18}$$

so that the rms displacement of the mapping, in grid units, is $d_{rms} = \delta n$. In practice (in the computer code) an unnormalized Gaussian field $\delta_u(\mathbf{k})$ is placed on the grid and we way normalize with a constant,

$$\delta_0(\mathbf{k}) = 2\pi \left[\sum_{\mathbf{k}} \frac{|\delta_u(\mathbf{k})|^2}{k^2} \right]^{-1/2} \delta_u(\mathbf{k}) \tag{4.19}$$

A normalized $\mathbf{u}(\mathbf{k})$ on the grid is

$$\mathbf{u}(\mathbf{k}) = n \left[\sum_{\mathbf{k}} \frac{|\delta_u(\mathbf{k})|^2}{k^2} \right]^{-1/2} i \frac{\mathbf{k}}{k^2} \tilde{\delta}_u(\mathbf{k}) \tag{4.20}$$

and the normalized k-space deformations (the FFT of Φ_{ij}) are given by

$$\Phi_{ij}(\mathbf{k}) = \frac{k_i k_j}{k^2} \delta_0(\mathbf{k}) \tag{4.21}$$

The above relations were written in three dimensions and apply equally well (with slight modifications) in two or one. For an unnormalized initial power spectrum $P_u(k)$, used to specify $\delta_u(\mathbf{k})$ on the grid, the above sums over $|\delta_u(\mathbf{k})|^2$ may be treated as integrals over $P_u(k)$, with the $\sum_{\mathbf{k}}$ going over to

$$\int 4\pi k^2 dk \qquad \int 2\pi k \, dk \qquad \int 2 \, dk \,, \tag{4.22}$$

in three, two, and one dimensional simulations. The k here is still referring to the discrete k_{int} values on a grid.

To generate a realization $\delta_0(\mathbf{k})$ of the Gaussian random field, we utilize a standard pseudo-random generator. From uniform deviates distributed between 0 and 1 we may transform to obtain Gaussian deviates of variance $P_0(k)/2$ for the real and imaginary parts of $\tilde{\delta}_0(\mathbf{k})$. For two uniform deviates r_1 and r_2, we generate

$$\delta_0(\mathbf{k}) = \sqrt{-\ln r_1 \, P_0(k)} \, e^{i2\pi r_2} \tag{4.23}$$

at each point **k** on half of the grid. In a series of simulations, by using the same initial seed in the pseudo-random sequence which covers k-space, we may generate realizations with the same initial phases but differing power spectra. Likewise, we may obtain multiple realizations of the same initial power spectrum, if each simulation starts with a different seed.

4.1.2 Linear and Nonlinear Scales

We now show the evolution from the linear to the nonlinear regimes. We first show the transition to nonlinear fluctuations in two dimensions, simply because the visual effect is clearest. Simulations in one and three dimensions follow. Although 3-d simulations have the least resolution—typically a 128^3 point grid—they will be emphasized the most throughout, since the universe is three-dimensional.

The linearly evolved rms local density contrast σ_l at the grid scale is

$$\sigma_l = \langle \delta^2 \rangle^{1/2} = b(t) \left(\int P_0(k)\, d^3k \right)^{1/2} \tag{4.24}$$

Note that here, and below, we may substitute d^dk for d^3k if needed, as in equation (4.22). Let us use the condition $\delta \approx 1$ to specify the onset of nonlinearity at a length scale k_{nl}.[1] This leads to the following definition for k_{nl},

$$b^2(t) \int_0^{k_{nl}} P_0(k)\, d^3k = 1 \tag{4.25}$$

The lower limit 0 here could in principle refer to a comoving horizon scale; practically, the fundamental mode $k = 1$ defines the lower limit in the sums over **k** modes. Small scales $k \gtrsim k_{nl}$ have gone nonlinear at time $b(t)$ given by (4.25). The nonlinear scale k_{nl} is a useful measure time evolution, in that k_{nl} decreases with time, i.e. as time goes on, larger and larger scales become nonlinear. The linear $b^2(t)P_0(k)$ time evolution of the power spectrum will hold for $k \lesssim k_{nl}$.

The nature of the Zel'dovich approximation and aspects of hierarchical clustering accessible by high k filtering are illustrated in Figures 4-1 through 4-5, which show the matter distribution for a series of two-dimensional simulations. The initial power spectrum is $P_i \sim k$ with a sharp cutoff at k_c.

We use a cutoff $k_c = 10$ and show the matter distribution at five epochs, defined by $\sigma_l = 0.1, 0.5, 1, 1.5$ and 2 according to (4.24). Statistical signals may be detected for data with $\sigma_l < 0.1$, but we do not display such data since they appear nearly uniform to the eye. These data are shown in the top panels of Figures 4-1 through 4-5. The evolution from the grid to a hint of structure is shown in the first two figures. Figure 4-3 shows the typical Zel'dovich approximation extrapolation, as we have $k_{nl} = k_c$. The last two show the approximation pushed a bit too far. In the lower panels, the same five $b(t)$ times are used, but the filtering has been performed at the nonlinear scale. The five k_{nl} values are 100, 20, 10, 7, and 5. At each successive epoch, larger scales are nonlinear. The σ_l, $b(t)$, and k_{nl} values for the fixed filter $k_c = 10$ sequence are provided in Table 4-1. The corresponding evolution parameters in the variable filter $k_c = k_{nl}$ sequence are shown in Table 4-2.

The k_{nl} parameter indicates the degree of evolution. For simulations with different initial power spectra, we evolve to the same k_{nl} in order to have analogous evolution. The meaning of k_{nl} is more robust in the case of full N-body simulations, since high-k filtering is not required as it is in the Zel'dovich approximation. Yet we still find it to be useful in our simulations. Note that the k_{nl} values in Table 4-1 are not the same as those in Table 4-2 even though both tables describe $P_i \sim k$ spectra; this is simply because the six different pre-filters we use in Table 4-2 amount to six different spectra. Specifically, our normalized initial spectra P_i are obtained from unnormalized $P_u(k)$, as can be seen in equation (4.19), by

$$P_i = C\,P_u(k) = \frac{(2\pi)^2}{\int P_u(k)\,k^{-2}\,d^3k}\,P_u(k) \tag{4.26}$$

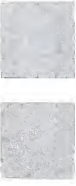


Figure 4-1 Two dimensional evolution, part 1 This is the first in a series of five plots (4-1–4-5) which show evolution from an initial $P_0 \sim k$ spectrum The top plots have a fixed cutoff $k_c = 10$, while the bottom plots use $k_c = k_{nl}$ (cutoff at the nonlinear scale) Top $\sigma_l = 0.1$ Bottom. $k_{nl} = 100$

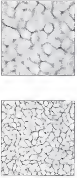


Figure 4-2 Two dimensional evolution part 2 Top $\sigma_l = 0.5$ Bottom $k_{ml} = 20$

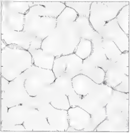


Figure 4-3 Two dimensional evolution part 3 $\sigma_l = 1$ and $k_{nl} = 10$

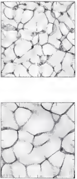


Figure 4-4 Two dimensional evolution part4 Top $\sigma_l = 1.5$ Bottom $k_{nl} = 7$

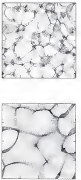


Figure 4-5 Two dimensional evolution part 5 Top $\sigma_l = 2$ Bottom $k_{nl} = 5$

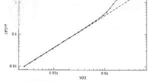


Figure 4-6 Growth of unit cell density fluctuations with time, in two dimensions. The dashed line is the 2-d linear perturbation theory result. The solid line and data points are the measured $\langle\delta^2\rangle^{1/2}$ results from a 2-d simulation.

Table 4-1 Linear evolution parameters for $P_0 \sim k^1$ with $k_c = 10$ in 2-d

σ_l	$b(t)$	k_{nl}
0.01	2.62×10^{-4}	215
0.1	2.62×10^{-3}	46
0.5	0.0131	16
1	0.0262	10
1.5	0.0392	7.6
2	0.0523	6.3

Table 4-2 Linear evolution parameters for $P_0 \sim k^1$ with $k_c = k_{nl}$ in 2-d

σ_l	$b(t)$	k_{nl}
1	2.62×10^{-4}	> 256
1	2.62×10^{-3}	100
1	0.0131	20
1	0.0262	10
1	0.0392	7
1	0.0523	5

where $P_u(k)$ has been filtered. We use isotropic pre-filters, $\delta(\mathbf{k}) \mapsto \delta(\mathbf{k})W(k/k_f)$, with $W(k/k_f) = 1$ for small k. The $P_u(k)$ therefore already includes a $|W|^2$ factor. In the present instance the spectra differ only in their sharp cutoff filtering scale k_f, but that is a significant difference, as can be seen in the figures. Choosing $k_f = k_{nl}$ is natural but not strictly required. If $k_f \lesssim k_{nl}$, then in general the Zel'dovich approximation is pushed too far, as shell crossing begins to wipe out the clustering. If $k_f \gtrsim k_{nl}$, then the approximation is well within the single-stream region and there is not much clustering. We take the P_0 in the integrand of equation (4.25) to be $CP_u(k)$ but without any filter, though this last distinction is only relevant in the cases where $k_f < k_{nl}$ (slight evolution). For instance, it allows us to take the scaling for $P_0 \sim k^n$ (in d dimensions)

$$\delta^2(t) \sim k_{nl}^{-(n+d)} \tag{4.27}$$

to be valid for various k_f/k_{nl} choices. For our fully evolved simulations, we take $k_f \approx k_{nl}$.

The lower sequence of plots in Figures 4-1 through 4-5 shows a nontrivial hierarchical evolution of structure formation. Full N-body simulations do not need initial high-k filtering. The fortunate fact is that the N-body results at successive epochs correlate well with k_{nl} filtered Zel'dovich approximation matter distributions, provided the final clumpy N-body distribution is smoothed (Coles et al. 1993). The correlations are more favorable for spectral index values $n = 0, -1$, and -2 which contain more initial small-k power. Interestingly, there is a piece of empirical evidence which helps explain the perhaps unexpected usefulness of a hierarchical pancaking model. The fully evolved N-body gravitational potential φ is found to resemble the k_{nl} smoothed initial φ_0 more than it resembles the unfiltered φ_0 (Pauls & Melott 1995). This at least indicates how the Zel'dovich approximation moves matter to the right place, overall.

The transition from the linear to the nonlinear regime is displayed in Figure 4-6 where we show the increase of the local rms fluctuations with time. The dashed curve in Figure 4-6 is σ_l vs. $b(t)$ as given by equation (4.24). The unit cell fluctuations, measured from a sequence of $P_0 \sim k$ simulations on a 512^2 grid, agree with linear theory early on, as they must. We use cloud-in-cell weighting to approximate a fine matter distribution on the simulation grid. On a 512^2 grid, using $k_c = 10$ on our spectrum, the mean displacement d_{rms} is only ≈ 1.3 unit cells when $\sigma_l = 0.1$, for example, so there are still some vestiges of grid effects prior to that. The nonlinearities inherent in the Zel'dovich approximation (mode coupling) cause the disagreement with linear theory beginning at $\sigma_l \approx 0.3$. The peak (and particularly the decrease after the peak) in Figure 4-6 illustrates the breakdown (shell crossing) of the approximation. The criterion $\sigma_l = 1$ for a fully extrapolated Zel'dovich approximation is simple yet effective. The farthest one could reasonably push the approximation, corresponding to the peak, is $\sigma_l = 1.2$ in this instance.

Similar considerations hold in the case of one-dimensional simulations. We illustrate a spectrum $P_0 = k^2$ with a cutoff $k_c = 10$ on a 32,768 cell 1-d grid. The hierarchical effect is shown in higher resolution in Figures 4-7–4-9. The top $1+\delta$ vs. x plots show the $\sigma_l = 0.1$, 1, and 2 evolution (with $k_c = 10$). The bottom plots are fully evolved with the cutoff at k_{nl}, using the same three $b(t)$ values. The unit cell rms fluctuations shown in Figure 4-10 agree with the $\sigma_l \sim b(t)$ 1-d perturbation theory early on, and the nonlinearities show up for $\langle\delta^2\rangle^{1/2} \approx 0.1$–1, but this local statistic gets a bit sloppy after that due to shell crossing. However, $\langle\delta^2\rangle^{1/2}$ does continue to increase with $b(t)$ until slightly past $\sigma_l = 1$, and such a nonlinearity criterion still proves useful. Incidentally, a double hump feature at shell crossing is a typical effect in 1-d simulations (using a finer grid or averaging over multiple realizations generally will not remove such an effect, but smoothing will remove it).

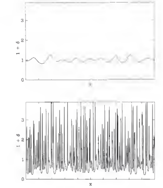


Figure 4-7 One dimensional evolution, part 1 This is the first of three plots (4-7–4-9) which show evolution from an initial $P_0 \sim k^2$ spectrum The top plots have a fixed cutoff $k_c = 10$ while the bottom plots use $k_c = k_{nl}$ (cutoff at the nonlinear scale) Top $\sigma_l = 0\ 1$ Bottom $k_{nl} = 100$

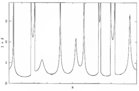


Figure 4-8 One dimensional evolution part 2 $\sigma_l = 1$ and $k_{nl} = k_c = 10$

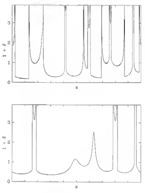


Figure 4-9 One dimensional evolution part 3 Top $\sigma_\ell = 2$ Bottom $k_{nl} = 5$

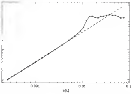


Figure 4-10 Growth of unit cell density fluctuations with time in one dimension. The dashed line is the 1-d linear perturbation theory result. The solid line and data points are the measured $\langle\delta^2\rangle^{1/2}$ results from a 1-d simulation.

We are obliged to focus on three dimensional simulations. On a 128^3 cell grid, we begin with $P_0 \sim k^{-1}$ correlations and a cutoff at $k_c = 10$. The evolution of structure is shown in Figures 4-11–4-13, which show slices from a sequence of 3-d simulations. In each case, the slice thickness is $L/64$ (where L is the cube dimension). The evolution of fluctuations with time is shown in Figure 4-14.

There is more large-scale power visible in the slices in Figures 4-11–4-13 than in the one and two dimensional simulations shown above. The primary reason is simply that we have chosen to illustrate a $P_0 \sim k^{-1}$ spectrum, which decreases with k. The secondary reason is that we are using a 2-d slice as our means of visualization. The density field covers an entire cube, yet we cut through a thin slice for our plots. Three dimensional structures forming perpendicular to the slice, e.g. the plane of a pancake coinciding with the plane of the slice, augment the power measured in two dimensions. The extra large-scale power is not always visibly clear in one-dimensional pencil beam sub-samples or two-dimensional slice sub-samples of a three dimensional field. However, we find that the extra small k power in estimators confined to 1-d or 2-d is noticeable compared to a 3-d $P(k)$ estimator.

These projection effects can be estimated as follows. We have a full three dimensional realization with density field $\delta(\mathbf{x})$, and we excise a thin slice in the xy plane as our sub-sample, with the 2-d density denoted by $\delta'(\mathbf{x})$. The power in the sub-sample is defined by

$$\langle \delta'(\mathbf{q}_1)\bar{\delta}'(\mathbf{q}_2)\rangle_{2d} = \left[\frac{(2\pi)^2}{L^2}\delta_D^{(2)}(\mathbf{q}_1+\mathbf{q}_2)\right]P_{2d}(\mathbf{q}_1), \tag{4.28}$$

where the 2-d wavevectors are denoted by $\mathbf{q}$. If we compute the left hand side of (4.28), confining $\mathbf{x}$ to the $z = 0$ plane, we find that $P_{2d}(q)$ is obtained directly from the full 3-d $P(k)$, but that $\mathbf{k}$ vectors perpendicular to the slice are contributing. Thus

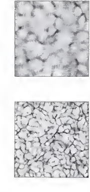


Figure 4-11 Three dimensional evolution part 1 This is the first of three plots (4-11 4-13) which show evolution from an initial $P_{\parallel} \sim k^{-1}$ spectrum Each plot shows a slice of thickness $L/64$ from a simulation on a 128^3 grid The top plots have a fixed cutoff $k_c = 10$, while the bottom plots use $k_c = k_{nl}$ (cutoff at the nonlinear scale) Top, $\sigma_l = 0.5$ Bottom $k_{nl} = 23$

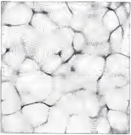


Figure 4-12 Three dimensional evolution part 2 $\sigma_l = 1$ and $k_{nl} = 10$

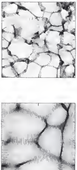


Figure 4-13 Three dimensional evolution part 3 Top: $\sigma_I = 2$ Bottom: $k_{rd} = 4$

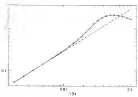


Figure 4-14 Growth of unit cell density fluctuations with time. The dashed line is the linear perturbation theory result. The solid line and data points are the measured $\langle\delta^2\rangle^{1/2}$ results from a 3-d simulation

we have

$$
\begin{aligned}
P_{2d}(q) &= \int \frac{L\,dk_z}{2\pi} P[(q^2 + k_z^2)^{1/2}] \\
&= \int_q^\infty \frac{L\,dk}{2\pi} \frac{k}{(k^2 - q^2)^{1/2}} P(k)
\end{aligned}
\tag{4.29}
$$

While on the topic of sub-samples, we go on to the case of a one dimensional pencil beam (which we take to be on the line $y = z = 0$). In analogy with the above case, we define

$$
\langle \tilde{\delta}'(\mathbf{q}_1) \tilde{\delta}'(\mathbf{q}_2) \rangle_{1d} = \left[\frac{2\pi}{L} \delta_D^{(1)}(\mathbf{q}_1 + \mathbf{q}_2) \right] P_{1d}(q_1), \tag{4.30}
$$

where 1-d wavevectors are denoted by $\mathbf{q}$. If we take a 1-d sub-sample from a 2-d realization of a density field with known power P_{2d}, we obtain

$$
P_{1d}(q) = \int \frac{L\,dk_y}{2\pi} P_{2d}[(q^2 + k_y^2)^{1/2}] \tag{4.31}
$$

More relevantly, we compute the left hand side of (4.30), assuming a 1-d sub-sample of a full 3-d realization, and we obtain

$$
\begin{aligned}
P_{1d}(q) &= \int \frac{L\,dk_y}{2\pi} \frac{L\,dk_z}{2\pi} P[(q^2 + k_y^2 + k_z^2)^{1/2}] \\
&= 2\pi \int_q^\infty \frac{L^2\,dk}{(2\pi)^2} k\,P(k)
\end{aligned}
\tag{4.32}
$$

These slice and pencil effects can be viewed as k-space projections in the following sense. In the case of the slice, we have a 3-d $\delta(\mathbf{x})$ field, and we multiply it by a function which is zero but for a thin slice near $z = 0$. In k-space, the $\tilde{\delta}(\mathbf{k})$ field is then convolved along k_z with a broad function. The above formula are simple since we multiplied $\delta(\mathbf{x})$ by a delta function (which has 1 as its transform). If $W(z)$ or $W(z, y)$ are the functions which produce a slice or pencil of finite thickness, then the $P(k)$ integrals in (4.29) and (4.32) would change according to

$$
P(k)\,dk \mapsto P(k) \left| \tilde{W}[(k^2 - q^2)^{1/2}] \right|^2 dk \tag{4.33}
$$

where, for example, we have

$$W(k_z) = \frac{\sin(k_z d/2)}{k_z d/2} \tag{4.34}$$

for a slice of thickness d in the xy plane.

Actual redshift surveys are often in the form of beams or slices, so these issues are relevant in practice (Kaiser & Peacock 1991), so as to not overestimate large-scale power. When implementing Lagrangian approximations, we find that it is best to use full 3-d simulations and then excise sub-samples if needed. While it is true that the properties of the initial gravitational potential are closely tied to the final caustic locations in one, two, or three dimensions, we find that the difference in the evolution of clustering between the 1-d, 2-d, and 3-d cases indicates that 3-d simulations are the way to go. This is somewhat obvious, but it needed to be investigated given the potential convenience of simulating P_{1d} and P_{2d} spectra directly. Anyway, our main focus is on full 3-d samples, and the slices we show are simply a helpful visual tool.

To investigate how far one may usefully push the approximation, we measured the local fluctuations $\sigma = \langle\delta^2\rangle^{1/2}$ versus $b(t)$ for various initial spectra in three dimensions. For $P_0 \sim k^n$, $-3 \leq n \leq 1$, we show the evolution of the local fluctuations in Figure 4-15. The linear theory result is also shown for the $n = -3$ case, plotted using dashed curves. The maximum evolution which could possibly be useful corresponds to the peaking of the curve at b_{max}, with $\sigma(b_{max}) = 2.6$ for $n = -3$ and $\sigma(b_{max}) = 2.0$ for $n = 1$. These maximum points are marked with asterisks. While perturbative results are in force, one would expect the σ vs. $b(t)$ curves to possess a positive second derivative $d^2\sigma/db^2$, since a perturbative expansion would necessarily involve powers $\sigma \sim b^p$ (beyond linear order $p = 1$). The σ vs. b inflection points are marked with triangles. Inflection points in σ^2 vs. b could also be used, with similar results (such points occur slightly later). The simplest criterion, based on extrapolating linear to its upper limit is $\sigma_l = 1$, and these points are plotted with solid squares in Figure 4-15. Using sharp k_c filtering in the Zel'dovich approximation, choosing $\sigma_l = 1$ is the same as setting $k_c = k_{nl}$. The curves in Figure 4-15 display a scaling according to $\sigma_l \sim b(t)$. This is shown in Figure 4-16, where we plot σ vs. σ_l for the same five data sets. Note that for the scaling to be accurately portrayed in Figure 4-16, one must compute the integrals used in the definitions of $P_0(k)$ and σ_l as discrete sums $\Sigma_{\mathbf{k}}$ rather than integrals $\int 4\pi k^2 dk$. Using integral expressions is generally fine, leading to errors on the order of a percent, but especially for indices $n < 0$, discrepancies will show up in direct comparisons with simulations.

The Zel'dovich approximation can be pushed to rms fluctuation values of $\sigma \approx 2$-3 before the approximation breaks down via shell crossing in the simulations. The high density structures (depending on the grid resolution) can reach $\delta_{max} \approx 10^1$–$10^2$. Such a dynamic range is smaller than that of a full N-body simulation, but it provides a useful probe into the nonlinear regime under controlled circumstances. Also, early epochs (where all scales remain in the linear regime) may be simulated so that we may witness the transition to nonlinearity. Generally, in the remaining pictures shown in this chapter and in the statistical analyses of Chapter 5, we concentrate on fully evolved Zel'dovich simulations. We let the scale of filtering k_f coincide with the scale of nonlinearity k_{nl} and push the approximation to its limits. Even then, the scales with $r \gtrsim k_{nl}^{-1}$ exhibit linear behavior, and the quasi-linear regime (which bridges the linear and highly nonlinear regimes) may be investigated.

4.2 Power and Bias

We now illustrate two of the most basic inputs into any cosmological simulation. First, there is the initial power spectrum $P_0(k)$, which completely specifies the initial

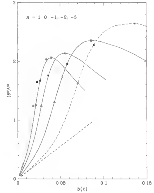


Figure 4-15 Measured growth of density fluctuations for various power indices in 3-d simulations The leftmost (dotted) curve corresponds to $n = 1$ Successive curves to the right apply to $n = 0$ -1 -2 -3 For $n = -3$ (dashed curve) we also plot the linear theory result

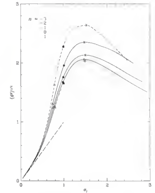


Figure 4-16 Measured growth of density fluctuations scaled to σ_l. The same data used in Figure 4-15 are plotted using $\sigma_l \sim b$. The linear perturbation theory result $\sigma = \sigma_l$ (dashed line) is also shown.

spatial correlations for the Gaussian field $\delta_0(\mathbf{x})$. Secondly, there is the biasing prescription which determines the nature of the galaxy sample to be analyzed. If we vary the bias or the initial power between two simulations, sometimes the distinction between them is visually clear. On the other hand, sometimes two different simulations look about the same and, though different, are difficult to distinguish clearly due to noise in the statistical measures. Most importantly, there are cases where relatively small changes in the initial parameters are clearly recognizable in the final statistical estimators. The 3-point function (presented in the following chapter) is an important example of a potentially discriminating statistic.

We assume that we have an evolved matter distribution, with density $\rho = \bar{\rho}(1+\delta)$ and that, based on this, there is a galaxy distribution. The number density of the galaxy distribution is given by $n_g = \bar{n}(1+\delta_g)$, and we must assume that there is at least some functional relationship between $\delta(\mathbf{x})$ and $\delta_g(\mathbf{x})$. As an approximate model we use a local bias,

$$\delta_g(\mathbf{x}) = f(\delta(\mathbf{x})) \tag{4.35}$$

Even if the function f is not known, δ and δ_g might be close enough after smoothing that a few terms in a Taylor expansion

$$\delta_g = f(\delta) = \sum_{n=0}^{\infty} \frac{1}{n!} b_n \delta^n \tag{4.36}$$

will suffice. We have $\langle\delta_g\rangle = \langle\delta\rangle = 0$ and b_0 is used to insure this zero average. b_0 plays no role in the statistical moments. The simplest case is linear bias, where $\delta_g = b_1\delta$ and such an effect is straightforward to trace through perturbation theory results. For example we have $P(k) \mapsto b_1^2 P(k)$. The linear $\mathbf{v}$ field in terms of δ_g, from equations (2.103) and (2.109) in §2.2.4, is

$$\mathbf{v}(\mathbf{x}) \approx \frac{\Omega^{0.6}}{b_1} \frac{H}{4\pi} \int d^3x' \delta_g(\mathbf{x}') \frac{\mathbf{x}' - \mathbf{x}}{|\mathbf{x}' - \mathbf{x}|^3} \tag{4.37}$$

which contains the typical combination $\beta = \Omega^{0.6}/b_1$ of the Ω and b_1 parameters. For example, averaging over a spherical overdensity of radius R (as in early Virgo infall estimates), we have an infall velocity v given by

$$\frac{v}{H_0 R} = \frac{1}{3}\frac{\Omega_0^{0.6}}{b_1}\langle\delta\rangle_R \tag{4.38}$$

A sharp threshold bias, where δ is clipped to zero for values below δ_t, is an example of a nonlinear biasing procedure. For an underlying Gaussian field, the the threshold model has been investigated analytically (Kaiser 1984, Politzer & Wise 1984, Bardeen et al. 1986, Jensen & Szalay 1986). For underlying local rms fluctuations σ_0 and threshold $\delta_t = \nu\sigma_0$, Kaiser (1984) found, for $\nu > 1$ and $\sigma_0 < 1$, a biased correlation function $\xi_t = b_\nu^2\xi/\sigma_0^2$ and $b_\nu = \nu$. Jensen and Szalay (1986) extended this result for b_ν to include small ν

$$b_\nu = \frac{\nu}{F(\nu)} \qquad F(\nu) = \sqrt{\frac{\pi}{2}}\,\nu e^{\nu^2/2}\,\mathrm{erfc}(\nu/\sqrt{2}) \tag{4.39}$$

where $F \to 1$ for large ν and $F \to \sqrt{\pi/2}\,\nu$ for small ν. A b_ν factor acts as a linear bias parameter b_1, and one might expect measured power to be amplified by such a b_1^2 factor in simulations, though our evolved density fields are not Gaussian (we find this to be so, qualitatively, in Chapter 5).

Going to second order in perturbation theory—e.g. as needed for the determination of the 3-point function—we need to include the first nonlinear bias term so that the biasing (4.36) is carried out to the same order as the perturbation theory. A modified hierarchical structure for the N-point functions persists even with such a bias carried out to all orders (Fry & Gaztañaga 1993). Keeping up to b_2 in the bias we have

$$\delta_g = b_0 + b_1\delta + \frac{b_2}{2}\delta^2, \tag{4.40}$$

where $b_0 = -b_2\langle\delta^2\rangle/2$. We show the biasing effect on the first few local density moments. Recall that the connected moments are

$$\langle\delta^2\rangle_c = \langle\delta^2\rangle \qquad \langle\delta^3\rangle_c = \langle\delta^3\rangle \qquad \langle\delta^4\rangle = \langle\delta^4\rangle_c + 3\langle\delta^2\rangle^2 \tag{4.41}$$

Using this order of bias, we have

$$\langle\delta_g^2\rangle = b_1^2\langle\delta^2\rangle + b_1 b_2\langle\delta^3\rangle + \frac{b_2^2}{2}\langle\delta^2\rangle^2 + \frac{b_2^2}{4}\langle\delta^4\rangle_c \tag{4.42}$$

where the second term shows the small effects of nonlinear bias on the second moment and the last two terms are actually superfluous unless one also includes a b_3 bias term. For the third moment, we find

$$\langle\delta_g^3\rangle = b_1^3\langle\delta^3\rangle + 3b_1^2 b_2\langle\delta^2\rangle^2 \tag{4.43}$$

The reduced hierarchical amplitude is thus

$$S_{3_g} = \frac{\langle\delta_g^3\rangle}{\langle\delta_g^2\rangle^2} = \frac{1}{b_1}S_3 + 3\frac{b_2}{b_1^2}, \tag{4.44}$$

to this order in perturbation theory and nonlinear bias. We utilize a similar effect on the Fourier space 3-point correlations in the next chapter. To get a handle on the biasing of our simulation data sets, we utilize these biasing parameters b_1 and b_2.

In practice, to implement a bias we impose a Poisson, or Monte Carlo, realization of the evolved density field. The simulated density field is already discrete, but it has as high a resolution as is practical; the resulting galaxy distribution is much more diluted or discrete. Without bias, galaxies are simply laid out with a probability proportional to the local value of $\delta(\mathbf{x})$, i.e. in practice we can take the discrete matter distribution from the simulation and simply dilute it uniformly randomly. For example, if there are $\approx 10^6$ simulation particles, we simply eliminate 99% of them randomly if we want $\approx 10^4$ galaxies in our simulated sample. Using this unbiased procedure, the galaxy statistics should be the same as the full $\delta(\mathbf{x})$ statistics, and while there is measurable signal, we find this to be true (if we account for discreteness effects). To implement bias, the Poisson procedure is of course explicitly dependent on the local $\delta(\mathbf{x})$ value. Favorite mechanisms include some sort of binary threshold, whereby a cell contains a galaxy if $\delta > \delta_c$ and a cell is empty otherwise, for some cutoff δ_c. One direct method is to keep or reject particles of the simulation according to some predefined probability $\mathcal{P}(\delta)$, for example we may a power law $\mathcal{P} \sim \delta^\alpha$.

Some insight into the effects of the initial P_0 spectrum can be obtained by inspecting slices from 3-d simulations, when there are variations in the initial power with a coherence scale corresponding to $k_c = 12$. Figures 4-17 through 4-23 show a set of realizations with the same initial phases but differing spectra. In each case we evolve the simulation fully, to $\sigma_l = 1$, corresponding to $k_{nl} = k_c$ since we truncate sharply. The slices are of thickness $L/64$, from 128^3 grid simulations. In each figure, we plot the linearly evolved spectrum $P = b^2(t)P_0$ below the slice. We begin by showing minimal $P_0 \sim k^4$ fluctuations, in Figure 4-17. A Zel'dovich $P_0 \sim k$ spectrum is used in Figure 4-18. There are strong cross correlations in terms of the major structures, though the measured 2 point statistics are clearly different. This trend continues in Figures 4-19 and 4-20, but there is more large-scale power, since the spectrum is a flat $P_0 =$ constant in the first case and constant from $k = 6$ to $k = 12$ (denoted $P_0 \sim$ hat(6 12) since this is a "top-hat" spectrum) in the second. Three negative power indices, $P_0 \sim k^n$ with $n = -1$, -2, and -3, are shown in Figures 4-21 through 4-23. Such a simulation does not correspond to a fair sample. Measured power from redshift surveys has turned over from $n = 1$ to $n \approx -1$, -2 on $50\,h^{-1}$ Mpc scales, so negative indices are quite relevant. Because the initial phases are the same, some of the same structures persist also in these figures.

Melott, Pellman, and Shandarin (1994) showed that the cross correlations between the Zel'dovich approximation and N-body simulations are improved if the k space

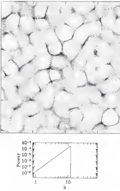


Figure 4-17 Power spectrum $P_0 \sim k^4$ sharp cutoff $k_c = 12$

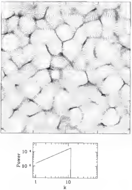


Figure 4-18 Power spectrum $P_2 \sim k$, sharp cutoff $k_c = 12$

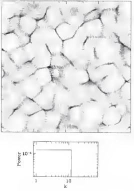


Figure 4-19 Power spectrum P_0 = constant sharp cutoff $k_c = 12$

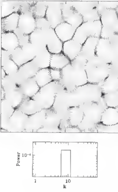


Figure 4-20 Power spectrum $P_k \sim \mathrm{hat}(6, 12)$

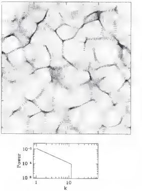


Figure 4-21 Power spectrum $P_0 \sim k^{-1}$ sharp cutoff $k_c = 12$

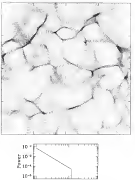


Figure 4-22 Power spectrum $P_0 \sim k^{-2}$ sharp cutoff $k_c = 12$

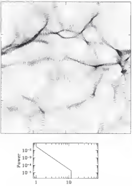


Figure 4-23 Power spectrum $P_0 \sim k^{-3}$ sharp cutoff $k_c = 12$

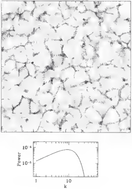


Figure 4-24 Power spectrum $P_0 \sim k$ Gaussian cutoff $k_G = 12$

Table 4-3 Evolution parameters for $P_0 \sim k^n$ with sharp k_{nl} cutoff

n	k_{nl}	$b(t)$	σ_l	σ
1	32	7.0×10^{-3}	1.0	0.96
1	16	0.014	1.0	1.4
1	8	0.028	1.0	1.8
1	4	0.055	1.0	2.1
0	32	8.5×10^{-3}	1.0	1.0
0	16	0.017	1.0	1.4
0	8	0.032	1.0	1.8
0	4	0.060	1.0	2.1
−1	32	0.013	1.0	1.1
−1	16	0.023	1.0	1.4
−1	8	0.041	1.0	1.8
−1	4	0.068	1.0	2.1
−2	32	0.028	1.0	1.1
−2	16	0.040	1.0	1.5
−2	8	0.056	1.0	1.8
−2	4	0.080	1.0	2.1

Table 4-4 Evolution parameters for $P_0 \sim k^n$ with Gaussian k_{nl} cutoff

n	k_{nl}	$b(t)$	σ_l	σ
1	32	7.0×10^{-3}	1.4	0.91
1	16	0.014	1.4	1.4
1	8	0.028	1.4	1.9
1	4	0.055	1.4	2.3
0	32	8.5×10^{-3}	1.4	1.0
0	16	0.017	1.5	1.4
0	8	0.032	1.5	1.9
0	4	0.060	1.5	2.3
−1	32	0.013	1.2	1.0
−1	16	0.023	1.3	1.5
−1	8	0.041	1.3	1.9
−1	4	0.068	1.3	2.3
−2	32	0.028	1.0	1.1
−2	16	0.040	1.1	1.4
−2	8	0.056	1.1	1.8
−2	4	0.080	1.1	2.2

truncation is performed with a Gaussian filter. The truncation idea is simply that

$$\delta_0(\mathbf{k}) \mapsto \delta_0(\mathbf{k})\tilde{W}(k, k_f) \tag{4.45}$$

for some isotropic filter which suppresses $k \gtrsim k_f$ power. The Gaussian filter is given by

$$W = \exp(-k^2/2k_G^2) \tag{4.46}$$

The simplest, yet effective (Coles et al. 1993), filter is the sharp step function $W(k, k_c) = 0$ for $k > k_c$. For any k_{nl} (Melott et al. (1994) found few differences between $k_{nl} = 4$ and $k_{nl} = 8$ for example) a good filter is $k_G \approx 1.25 k_{nl}$ for $P_0 \sim k^n$ with $n = -2, -1$, and 0, and $k_G \approx k_{nl}$ for $n = 1$. For example, using $P_0 \sim k$ and $k_{nl} = 12$, as in Figure 4-18, we choose a Gaussian window with $k_G = 12$ and plot the results in Figure 4-24. This correlates well with Figure 4-18 and with a full $P_0 \sim k$ (no cutoff) N-body simulation (after smoothing).

Simulation sequences using both sharp and Gaussian cutoffs have been performed for k^n spectra. For our k_{nl} filtering, we take $k_c = k_{nl}$ for the top hat filter. For the Gaussian case, we take $k_G = k_{nl}$ for $n = 1$ and $k_G = 1.25 k_{nl}$ for $n \leq 0$. For a given index n being k_{nl} filtered by either a top hat or a Gaussian, the evolution as marked by $b(t)$ is practically the same (the two $b(t)$ values differ by $\lesssim 1\%$ generally). The Gaussian case possesses more small scale power but also allows slightly more shell crossing. As shown by Melott et al. (1994), the compromise between these two effects works out nicely regarding N-body comparisons. In Table 4-3 we show evolution parameters for $P_0 \sim k^n$ spectra with sharp k_{nl} filtering. The k_{nl}, $b(t)$, and σ_l columns are linear theory parameters, whereas $\sigma = \langle\delta^2\rangle^{1/2}$ is the measured rms unit cell fluctuation. Table 4-4 shows the analogous values in the case of a Gaussian filter.

For statistical purposes, the matter distribution is characterized by the $\delta(\mathbf{x})$ array on the simulation grid. In practice, we have a counts-in-unit-cells array, cuc($\mathbf{x}$), evaluated at the 128^3 grid points using cloud-in-cell weighting (so that the counts need not be integers). For example, with $P_0 \sim k^{-1}$ and Gaussian filtering $k_G = 16$ for $k_{nl} = 12$, we extract a thin $L/128$ slice corresponding to one layer of the cic array. This is shown in Figure 4-25. In the gray scale used in the plot, we have clipped density contrast values above 5, although values over 30 existed in the slice plotted. We do this because underdense regions dominate the distribution (the median δ value is ≈ -0.5, and the δ distribution has positive skewness). It may appear a bit course, but recall that there are 2 097 152 elements in the full 3-d cic array. In the analyses, we may sharpen the measured spectra after the loss of resolution after cloud-in-cell binning (for the $\delta(\mathbf{x})$ field) or discrete, integer binning (for the $\delta_g(\mathbf{x})$ field) by deconvolution. This is generally not a large effect (cf. Chapter 5). Figure 4-26 shows galaxy distribution slices stemming from the $\delta(\mathbf{x})$ in Figure 4-25. Each slice (of thickness $L/64$) has $\approx 3\,000$ galaxies, as a subsample of $\approx 2 \times 10^5$ galaxies in the cube. Slice (a) is unbiased and has the same statistics as the undiluted $\delta(\mathbf{x})$, not including the discreteness shot noise. In slices (b), (c), and (d), we apply bias with a threshold δ_t and vary the bias probability $\mathcal{P}(\rho)$. In (b), $\delta_t = 1$ and we take $\mathcal{P} = 0.1$ for $\delta < \delta_t$. In (c), we have $\mathcal{P} \sim \rho^{1/2}$ below $\delta_t = 1$. In (d), we have $\mathcal{P} \sim \rho$ below $\delta_t = 3$. In Figure 4-27 we show two diluted samples with $\mathcal{P} \sim \rho^{1/2}$. For the top slice $\delta_t = 1$, the bottom has $\delta_t = 3$. The galaxies in the simulation cube have been diluted to $\approx 5 \times 10^4$, and each slice has ≈ 800 galaxies. Shot noise is significant here, but the measured 2-point and 3-point spectra can still be distinguished and, with the help of perturbation theory, be used to analyze the dynamics and biasing.

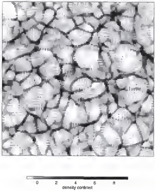
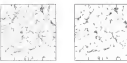
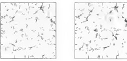


Figure 4-25 A $L/128$ thin slice from the 128^3 cic array corresponding to the unbiased $\delta(\mathbf{x})$ field Power spectrum $P_0 \sim k^{-1}$, Gaussian cutoff $k_G = 15$ $k_{nl} = 12$



(a) unbiased (b) $\mathcal{P}=0.1$ $\delta < 1$



(c) $\mathcal{P}\sim\rho^{1/2}$, $\delta < 1$ (d) $\mathcal{P}\sim\rho$, $\delta < 3$

Figure 4-26 Four galaxy distribution slices which are the same except for the biasing, based on the $\delta(\mathbf{x})$ in Figure 4-25 Plot (a) has no biasing Different biasing schemes, with bias probability $\mathcal{P}$ below a threshold δ_t are used for the other three plots

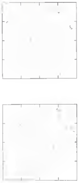


Figure 4-27 Two dilute samples derived from the matter distribution in Figure 4-25 Each is based on a bias probability $\mathcal{P} \sim \rho^{1/2}$ below a threshold δ_t Top $\delta_t = 1$ Bottom. $\delta_t = 3$

CHAPTER 5
STATISTICS

5.1 Introduction

The statistical analysis of galaxy surveys and simulated cosmological data is an important part of the large-scale structure problem. We focus on 2-point and 3-point spatial correlations primarily because they provide practical, measurable descriptions of observable matter distributions and because there are perturbative results which apply to large-scale data. In principle, a full set of N-point correlations would be needed to completely describe an evolved density field, but even the 2-point function can have a noisy signal, and higher order correlations are smaller and have decreasing signal to noise ratios. Recall that the 2-point function completely specifies the initial Gaussian density field, as all higher order connected moments are equal to zero. As the matter distribution evolves gravitationally, the power spectrum $P(k)$, the Fourier space 2-point function, shows a strong imprint of the initial spectrum in the linear regime and retains a noticeable imprint even as nonlinearities develop. The bispectrum $B(k_1, k_2, k_3)$, the 3-point function in Fourier space, is the first correlation function devoted solely to nonlinearities and proves to be a rich statistic. For example, the $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$ triangle shape dependence probes the inherently anisotropic nature of gravitational instability, and to lowest nonvanishing order in perturbation theory the effects of clustering and bias are largely separable (Fry 1994b). We concentrate on the power spectrum and bispectrum in analyzing the matter and (possibly biased) galaxy density contrast fields gathered from simulations similar to those we presented in Chapter 4.

The Zel'dovich approximation allows us to probe the mildly nonlinear regime of structure formation. This is also known as the quasi-linear regime and marks the transition between the linear and highly nonlinear stages of evolution.

At early times, the local rms density contrast is small $\sigma \ll 1$ and the density fluctuation modes evolve independently according to first order perturbation theory

$$\tilde{\delta}(\mathbf{k}, t) = b(t)\,\delta_0(\mathbf{k}) \tag{5.1}$$

for all $\mathbf{k}$. The local growth of fluctuations is well described by the first order relation $\sigma_l \sim b(t)$. At these early epochs, the linear theory applies to all scales, and $\delta(\mathbf{x})$ remains a Gaussian field.

After significant evolution we have $\sigma > 1$ and nonlinear evolution governs the smaller scales while linear and quasi-linear regimes exist on larger scales. In general, the rms density contrast σ_R filtered at length scale R is a decreasing function of R. At later epochs, the large scales for which $\sigma_R \ll 1$ remain in the linear regime, and small scales for which $\sigma_R > 1$ have gone nonlinear. The condition $\sigma_R \approx 1$ marks the transition. We identify epochs in our simulations by k_{nl}, the wavenumber which marks the nonlinear transition according to extrapolated linear theory:

$$b^2(t) \int_0^{k_{nl}} P_0(k)\, d^3k = 1 \tag{5.2}$$

In the quasi-linear regime (roughly $\sigma \lesssim 1$), the coupling of different modes is important, especially for $k \gtrsim k_{nl}$, and higher order correlations such as the bispectrum develop.

The main goal in section §5.2 is to measure 2-point and 3-point correlations in Fourier space. The data to be analyzed are the matter and galaxy density contrast fields $\delta(\mathbf{x})$ and $\delta_g(\mathbf{x})$ from Zel'dovich approximation simulations. We present initial results concerning the power spectrum in §5.2.1. The effect of mode coupling is illustrated in §5.2.2. In §5.2.3 we present the reduced bispectrum measurements. In

§5.3 we investigate the power spectrum and reduced bispectrum while changing the initial statistics and biasing procedures in the simulations. The main importance of this section lies in the fact that we are probing the reduced bispectrum within a controlled environment.

5.2 Measured Correlations

5.2.1 Power Spectrum

We present power spectrum results for the underlying matter distribution and for various discrete samples. The simulations are performed in a cubical comoving volume. This entire volume comprises our sample and is filled with an $n \times n \times n$ grid (typically $n = 128$). The cloud-in-cell binned unbiased $\delta(\mathbf{x})$ array represents the underlying matter distribution, which is practically continuous (i.e. discreteness effects are generally not important in this instance). The power spectrum is obtained from $\tilde{\delta}(\mathbf{k})$, the discrete Fourier Transform (FFT) of $\delta(\mathbf{x})$. The discrete "galaxy" samples are analyzed on the same grid, with the galaxy density contrast denoted by $\delta_g(\mathbf{x})$. Discreteness effects potentially are important for the galaxy distribution. No simulated selection effects are imposed (the samples are volume limited), and the cubical geometry is retained.

The raw (uncorrected) power is obtained by averaging $|\tilde{\delta}(\mathbf{k})|^2$ over all wavevectors of fixed length $k \pm \frac{1}{2}\Delta k$. Wavevectors are expressed as multiples of the fundamental mode, whose wavelength is L, the length of the cube. For one realization, the number of modes averaged over is approximately $2\pi k^2 \Delta k$. Since the grid size is $n \times n \times n$, our Nyquist frequency is $n/2$ (typically 64 since we use $n = 128$). The k space bin size for $P(k)$ measurements is taken to be $\Delta k = 1$. The error bars shown in measurements below are the statistical (standard deviation of the mean) 1-σ errors. To obtain the clustering power signal $P(k)$, we include discreteness and grid blurring corrections described below.

Results for initial power law spectra filtered at the nonlinear scale are shown in Figures 5-1 through 5-8. The measured power spectra at four stages of sharply k_{nl} filtered evolution are shown in Figures 5-1 through 5-4. The initial spectra are k^n power laws, and results for indices $n = 1$, 0, -1, and -2 are plotted in each figure. Figure 5-1 refers to the earliest stage, $k_{nl} = 32$, while k_{nl} values of 16, 8, and 4 in Figures 5-2 through 5-4 refer to increasingly later stages. Results from the analogous simulations which used Gaussian high-k filtering are shown in Figures 5-5 through 5-8. As in the previous chapter, we use $k_G = k_{nl}$ for $n = 1$ and $k_G = 1.25k_{nl}$ for $n \leq 0$.

The spectra shown in Figures 5-1 through 5-8 were obtained by averaging over one realization, and the signal is still relatively sharp. We have averaged over 10 realizations in the $k_c = 16$ case (Figure 5-2) to illustrate how such averaging reduces the statistical error bars and clarifies the signal. Below the nonlinear filtering scale ($k \lesssim k_{nl}$), the spectra generally follow the linear predictions. For $k \gtrsim 0.5k_{nl}$, the transfer of power to smaller scales is clearly shown. The linear normalization (obtained from the initial conditions) is approximately correct. Higher order corrections affect the shape of $P(k)$ in the linear regime, rather than merely changing the amplitude and leaking power to small scales. Figure 5-2 also shows the minor differences between using $\int d^3k$ or the strictly correct discrete sums $\sum_{\mathbf{k}}$ in computing the linear normalization. Recall that the local variance $\sigma^2 = \langle \delta^2 \rangle$ is the integral of the power spectrum: the integral of $k^2 P_0$ is σ_l^2, while the integral of measured P gives the measured σ. At small scales, e.g. $k \gtrsim 2k_{nl}$, we find that an effective power law $P \sim k^{-\beta}$ emerges over a range in k which approaches the Nyquist limit. The power law is approximately k^{-3}. The least squares fits typically yield $\beta = 2.6$–3.0, independent

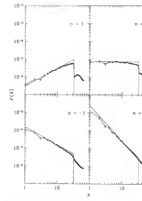


Figure 5-1 Measured power for initial $P_0 \sim k^n$ with sharp cutoff $k_c = k_{nl} = 32$. The linear power is also plotted

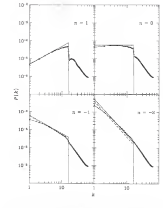


Figure 5-2 Measured power for initial $P_0 \sim k^n$; $k_c = k_{nl} = 16$. The power was averaged over 10 realizations to clarify the signal. The linear power is also plotted (the solid curves from $\int d^3k$ computations, the dashed using the strictly correct discrete sums $\sum_{\mathbf{k}}$)

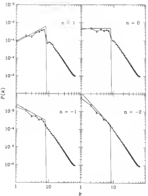


Figure 5-3 Measured power for initial $P_0 \sim k^n$ $k_c = k_{nl} = 8$. The linear power is also plotted

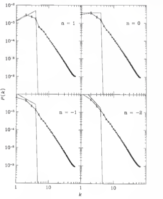


Figure 5 4 Measured power for initial $P_0 \sim k^n$ $k_c = k_{nl} = 4$ The linear power is also plotted

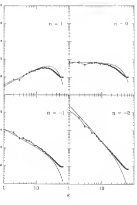


Figure 5-5 Measured power for initial $P_0 \sim k^n$ with Gaussian cutoff $k_G \approx k_{nl} = 32$ The linear power is also plotted

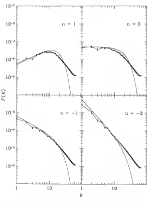


Figure 5-6 Measured power for initial $P_0 \sim k^n$; $k_G \approx k_{nl} = 16$. The linear power is also plotted

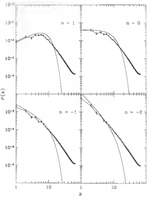


Figure 5-7 Measured power for initial $P_0 \sim k^n$ $k_G \approx k_{nl} = 8$ The linear power is also plotted

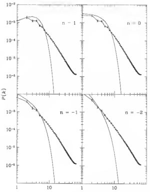


Figure 5-8 Measured power for initial $P_0 \sim k^n$ $k_G \approx k_{nl} = 4$ The linear power is also plotted

of the initial k^n index. For example the 8 runs with $k_{nl} = 4$ (4 different indices n, 2 different filters) give $\beta = 2.89 \pm 0.04$ fitting from $k = 16$–48. Fitting over the same k range in the $k_{nl} = 8$ case gives $\beta = 2.72 \pm 0.03$.

The power spectrum was measured at various times for fixed initial spectra. The general expectation is that $P \sim b^2(t)$ for $k \lesssim k_{nl}$. For k^1 and k^{-1} initial spectra with a sharp cutoff $k_c = 8$ we plot the power spectrum in Figures 5-9 and 5-10 for six stages of evolution $\sigma_l = 0.125$, 0.25, 0.5, 1.0, 2.0 and 4.0. For $k < k_c$ the $b^2(t)$ scaling holds even to the unreasonably high $\sigma_l = 4$ epoch. For $k > k_c$, $P \sim b^4(t)$ seems to be the dominant effect early on. This is shown in Figure 5-11, where P versus σ_l is plotted for fixed $k = 2$ and $k = 12$.

The discreteness or shot noise contribution to the power is $1/N$ for N "galaxies." Taking this into account, the raw power per mode signal is related to the clustering 2-point signal $P(k)$ by

$$\langle|\delta(\mathbf{k})|^2\rangle = P(k) + \frac{1}{N} \tag{5.3}$$

This follows by analyzing the FFT sum used to generate $\delta(\mathbf{k})$,

$$\delta(\mathbf{k}) = \frac{1}{V}\sum_{\mathbf{x}} \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{5.4}$$

where $V = L^3 = n^3$ is the volume. Using a grid counts array $m(\mathbf{x})$

$$\delta(\mathbf{x}) = \frac{m(\mathbf{x})}{\bar{m}} - 1 \qquad \bar{m} = \frac{N}{V} \tag{5.5}$$

we have

$$\delta(\mathbf{k}) = -[\delta_D(\mathbf{k})] + \frac{1}{V}\sum_{\mathbf{x}} \frac{m(\mathbf{x})}{\bar{m}} e^{-i\mathbf{k}\cdot\mathbf{x}} \tag{5.6}$$

which for $\mathbf{k} \neq 0$ is just

$$\delta(\mathbf{k}) = \frac{1}{N}\sum_{\mathbf{x}} m(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \tag{5.7}$$

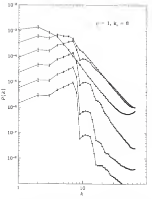


Figure 5-9 Power $P(k)$ vs. k at several epochs, for an initial spectrum $\sim k^1$. $k_c = 8$. The six times are $\sigma_l = 0.125, 0.25, 0.5, 1.0, 2.0$ and 4.0 corresponding to the six curves, bottom to top from the left.

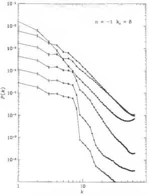


Figure 5-10 Power $P(k)$ vs k at several epochs for an initial spectrum $\sim k^{-1}$ $k_c = 8$ The six times are $\sigma_l = 0.125$ 0.25 0.5 1.0 2.0 and 4.0 corresponding to the six curves bottom to top from the left

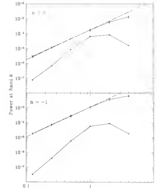


Figure 5-11 Power $P(k)$ at fixed k versus σ_l The initial spectrum is $\sim k^n$ $k_c = 8$ with $n = 1$ and $n = -1$ In each panel we plot P versus σ_l at $k = 2$ (top curve) and $k = 12$ (bottom curve) The dashed line has slope 2 the dotted line slope 4

If $m(\mathbf{x})$ is 0 or 1 and if $m(\mathbf{x})$ represents a uniform random process, then the sum (5.7) is a random walk of N steps of length $1/N$ in the complex plane, and the power is

$$\langle|\tilde{\delta}(\mathbf{k})|^2\rangle = \frac{1}{N} \tag{5.8}$$

Even with no clustering signal, this noise would be present. Discussions on discreteness effects can be found in Peebles (1980), Baumgart and Fry (1991), and Fry, Melott and Shandarin (1994).

Another natural candidate for discreteness corrections to the power would be $1/N_{occ}$, where N_{occ} is the number of occupied cells in the $m(\mathbf{x})$ array. If there were effectively N/N_{occ} galaxies per cluster on the scale of the grid cell, the random noise in the power signal of the clusters becomes $1/N_{occ}$. A third possibility for a correction follows if we do not assume that $\langle m^2\rangle = \langle m\rangle$ in the computation of the $\langle\tilde{\delta}(\mathbf{k})\tilde{\delta}(\mathbf{k}')\rangle$ sum, which leads to the following expression for the raw (uncorrected) power

$$\langle|\tilde{\delta}(\mathbf{k})|^2\rangle = \frac{\langle m^2\rangle V}{N^2} + \frac{1}{V}\sum_{\mathbf{x}} \xi(x) e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{5.9}$$

where ξ is the spatial 2-point function. In the second term, which is what we call the clustering power spectrum, we have $\langle m(\mathbf{x})m(\mathbf{x}')\rangle = \bar{m}^2(1+\xi(|\mathbf{x}-\mathbf{x}'|)$. The first term does reduce to $1/N$ and $1/N_{occ}$ in the respective limits.

To check the validity of the code which computes the $P(k)$ estimator, and to validate our ideas on $1/N$ noise, we analyze data which are generated by a uniform Poisson process. We place N points into the sample volume with a pseudo-random number generator and then bin them in the 128^3 unit cells. For one realization of $N = 1000$ and one of $N = 10000$, we have plotted the raw power spectrum in Figure 5-12. The dashed lines indicate the expected $1/N$ signal. Convergence on the expected value improves with multiple realizations. In the case of cloud-in-cell binning, which we use only for the unbiased underlying matter distribution, discreteness should not generally be relevant. Since $128^{-3} \approx 5 \times 10^{-7}$, such a signal could play a role if one is investigating a small clustering signal generated in a simulation (at early times for example) For the uniform Poisson $N = 128^3$ case we see in Figure 5-12 that the $1/N$ power is modulated by a $\mathrm{Sa}^4(\pi k/128)$ factor where

$$\mathrm{Sa}(y) = \frac{\sin(y)}{y} \tag{5 10}$$

The Sa^4 factor arises because $\mathrm{Sa}^2(\pi k/128)$ is the transform of the "triangle" function associated with smearing a cubical cloud over cubical cells The blurring due to grid binning is addressed below in the context of "sharpening" the signal

When the particles of the simulation are binned to the unit cells of the grid for the FFT resolution is lost This loss of resolution in $\delta(\mathbf{x})$ can be partly corrected for by sharpening the Fourier components as follows. If a cell of size a is used to bin the data then this is essentially equivalent to convolving $\delta(x)$ with a rectangle function in each dimension. The rectangle function $\mathrm{Rect}(x, a)$ is zero except for an interval of length a.

$$\mathrm{Rect}(x, a) = \begin{cases} 1/a & \text{if } |x| < a/2 \\ 0, & \text{otherwise} \end{cases} \tag{5 11}$$

The Fourier transform of $\mathrm{Rect}(x, a)$ is

$$\mathrm{Sa}(\frac{\pi k_x a}{L}) \tag{5 12}$$

where we use integer wavevectors for a box of length L In three dimensions, we may sharpen (deconvolve) the $\delta(\mathbf{k})$ array by dividing each element (in the case of a unit cell, $a = 1$) by

$$S(k_x, k_y\ k_z) = \mathrm{Sa}(\frac{\pi k_x}{L})\,\mathrm{Sa}(\frac{\pi k_y}{L})\,\mathrm{Sa}(\frac{\pi k_z}{L}) \tag{5 13}$$

The sharpening for cloud-in-cell binning is similar For cloud-in-cell binning, the cloud can span parts of two unit cells in each dimension The c-i-c binned data is like the ideal $\delta(\mathbf{x})$ data convolved with a triangle (with a base of length 2 and height 1/2) function in each dimension Note that such a triangle function is the convolution of

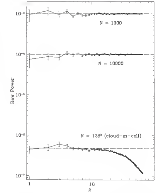


Figure 5.12 Raw power spectrum of a uniform Poisson process with N particles. The dashed lines show the $1/N$ values. In the c-i-c case, the points follow $1/N$ modulated by $\mathrm{Sa}^4(\pi k/128)$—the dotted curve that can be seen for $k > 50$ (where the fit worsens).

two Rect(x,a) functions, so that its transform is the square of (5.12). Thus we may sharpen by

$$S_{cic}(k_x, k_y, k_z) = \mathrm{Sa}^2(\frac{\pi k_x}{L})\mathrm{Sa}^2(\frac{\pi k_y}{L})\mathrm{Sa}^2(\frac{\pi k_z}{L}) \tag{5.14}$$

in this instance.

In practice, we wish to estimate the raw power and bispectrum from unsharpened data and apply discreteness corrections, because it is the unsharpened Poisson data which gives $1/N$ (or Sa^4/N for c-i-c). We may conveniently sharpen after the discreteness corrections have been made since the $S(k_x, k_y, k_z)$ functions defined above are so nearly isotropic. We use the approximation

$$\begin{aligned} S(k_x, k_y, k_z) &= 1 - \frac{\pi^2}{6L^2}(k_x^2 + k_y^2 + k_z^2) + \text{anisotropic} \\ &\approx \mathrm{Sa}(\frac{\pi k}{L}) \end{aligned} \tag{5.15}$$

which shows that the sharpening correction is roughly a function of k only. We plot the Sa factor (for the typical case $L = 128$) in Figure 5-13 with error bars showing the maximum difference between the isotropic and anisotropic functions. We apply sharpening corrections after the P and B estimates by applying the appropriate number of Sa factors. The k_{nl} values we choose are safely short of $k = L/2$ (the maximum k_{nl} value we choose is $L/4$).

We have examined the effect of dilution on correlation measurements. While varying N, the number of particles in the sample, the clustering power becomes lost in the noise for small N. Given the dynamic range of our simulations, with our typical level of evolution, e.g. $k_{nl} = 16$, an unbiased power signal cannot be usefully be resolved for $N \lesssim 10^3$. The degradation of the power signal for a $P_0 \sim k^{-1}$ simulation as the sample is diluted is shown in Figure 5-14. The signal is effectively a power law k^{α} for small $k < k_{nl}$ and k^{β} for $k > k_{nl}$. We fit over the ranges 1–10 and 20–40 and list the power indices in Table 5-1. For $N \approx 10^3$, only 44 of the 64 modes have raw power above the $1/N$ noise level.

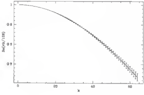


Figure 5-13 Sharpening correction. The error bars show the maximum difference between the isotropic approximation $\mathrm{Sa}(\pi k/128)$ and $\mathrm{Sharp}(k_x, k_y, k_z)$. Even at the Nyquist frequency ($k = 64$) this error is only 3%.

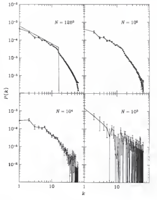


Figure 5-14 Degradation of power signal with dilution Four samples from a $P_0 \sim k^{-1}$ $k_c = k_{nl} = 16$ are taken The $N = 128^3$ panel shows the linear spectrum as well as the underlying cic binned and discretely binned result—only for $k > 60$ can the discretely binned power be seen to dip The $N = 10^3$ signal is not useful for $k > k_{nl}$

Table 5-1 Power law indices for dilute samples: k^{α} for $k \lesssim k_{nl}$, k^{β} for $k \gtrsim k_{nl}$

N	α	β
128^3 (c-i-c)	-1.00 ± 0.05	-2.85 ± 0.01
128^3	-1.00 ± 0.05	-2.75 ± 0.01
10^5	-0.99 ± 0.05	-2.58 ± 0.04
10^4	-0.98 ± 0.08	-2.46 ± 0.29
10^3	-1.48 ± 0.23	-0.9 ± 0.7

We use self consistency as the criterion for choosing a discreteness correction and the $1/N$ choice generally proves best. Experiments which compare the variance of the power signals as only N is varied tend to bear this out.

5.2.2 Mode Coupling

In §3.3.1, we obtained a perturbative expression for $\tilde{\delta}(\mathbf{k})$ which explicitly displayed the coupling of modes (Grinstein & Wise 1987),

$$\begin{aligned}\tilde{\delta}(\mathbf{k},t) &= \sum_{n=1}^{\infty} b^n \int d^3k_1 \cdots d^3k_n [\delta_D] G_n(\mathbf{k}_1, \dots, \mathbf{k}_n) \tilde{\delta}_0(\mathbf{k}_1) \cdots \tilde{\delta}_0(\mathbf{k}_n) \\ G_n(\mathbf{k}_1, \dots, \mathbf{k}_n) &= \frac{1}{n!} \frac{(\mathbf{k}\cdot\mathbf{k}_1)}{k_1^2} \frac{(\mathbf{k}\cdot\mathbf{k}_2)}{k_2^2} \cdots \frac{(\mathbf{k}\cdot\mathbf{k}_n)}{k_n^2}\end{aligned} \qquad (5.16)$$

where we set $\mathbf{k} = \sum \mathbf{k}_i$ due to the $[\delta_D]$ constraint. We apply results based on (5.16) to a few special cases below.

The power spectrum is given by

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle = [\delta_D(\mathbf{k} + \mathbf{k}')]\, P(k) \qquad (5.17)$$

The product $\langle \delta(\mathbf{k}) \tilde{\delta}(\mathbf{k}') \rangle$ is computed order by order, accounting for the scaling of the perturbative series, $\tilde{\delta}^{(n)} \sim (\tilde{\delta}^{(1)})^n$. We utilize the formalism of Fry (1984), Goroff et al. (1986) and Fry (1984a). Since $\tilde{\delta}^{(1)} = b(t)\tilde{\delta}_0$ is a Gaussian field, all its odd statistical moments vanish, so the first corrections to the $P(k)$ are $\sim (\tilde{\delta}^{(1)})^4$ and therefore involve second and third order perturbation theory. Explicitly factoring out the $b(t)$ dependence, we have

$$\begin{aligned} P(k) &= b^2 P_{11} + b^4 [2P_{13} + P_{22}] + b^6 [P_{15} + 2P_{24} + P_{33}] + \cdots \\ \langle \tilde{\delta}^{(m)} \tilde{\delta}^{(n)} \rangle &= [\delta_D]\, b^{m+n} P_{mn} \end{aligned} \tag{5.18}$$

Note that $P_{11} \equiv P_0$. Fry (1994a) computes terms in this series up to order b^{10} and applies it to order b^6, which utilizes fifth order theory, in the case of the exact dynamics. The first nonlinear corrections are

$$\begin{aligned} P_{13} &= 3P_0(k) \int d^3k\, G_3(\mathbf{k}, \mathbf{k}', -\mathbf{k}')\, P_0(k') \\ P_{22} &= 2 \int d^3k\, G_2^2(\mathbf{k}', \mathbf{k} - \mathbf{k}')\, P_0(k')\, P_0(|\mathbf{k} - \mathbf{k}'|) \end{aligned} \tag{5.19}$$

These two terms represent the 1-loop corrections to the 2-point correlations, as investigated by Scoccimarro and Frieman (1996), who extensively develop perturbative calculations beyond leading order (i.e. beyond the tree level contributions). See Fry (1984), Goroff et al. (1986), and Scoccimarro and Frieman (1996) for discussions of a diagrammatic approach to perturbative calculations of N-point correlations.

For the Zel'dovich kernel (5.16) we compute these corrections in the case of a top hat initial spectrum, $P_0 \sim \mathrm{hat}(k_1, k_2)$. The function $\mathrm{hat}(k_1, k_2)$ is 1 for $k_1 \le k \le k_2$ and 0 otherwise. Such a spectrum contains no initial power at large scales ($k < k_1$) or small scales ($k > k_2$) and so provides an environment to display mode coupling explicitly. A k^4 small k tail (Zel'dovich 1965) should be detectable, since there is no linear power to dominate it.

Zel'dovich (1965) and Peebles (1974b, 1980) describe how nonlinearities induce a $P \sim k^4$ spectrum on large scales when no large-scale power is initially present. Fry (1994a) shows how this behavior persists to higher orders in perturbation theory. In the case of a homogeneous distribution of particles, with interparticle length scale x_0, simply displacing the mass will generate a minimal power signal as $k \to 0$. Naively, by looking at the $\delta(\mathbf{k})$ transform (5.4), one would obtain $\tilde{\delta}(\mathbf{k}) \sim k x_0$ for small k, but in fact such a term vanishes due to momentum conservation. The next term gives $\delta(\mathbf{k}) \sim (kx_0)^2$, which supplies the minimal k^4 spectrum. In our 3-d c-i-c binned simulations, this signal is marginally observable (due to limited dynamic range) as we show below. Note that this effect is observed without c-i-c discreteness corrections, and such a signal can be detected far below the $1/N$ shot noise level. In 2-d or 1-d simulations, the effect is more easily observed.

We perform the 1-loop corrections for the case

$$P_0 = C\,\mathrm{hat}(k_1, 2k_1) \tag{5.20}$$

The results depend on the combination Cb^2, and we fix C by our $\int d^3k\, P_0(k) k^{-2}$ simulation normalization equation (4-19). This gives $C = \pi/k_1$, and the local linear $\langle\delta^2\rangle$ fluctuations are

$$\sigma_l = \int d^3k\, b^2(t) P_0(k) = \frac{28}{3}\pi^2 k_1^2 b^2 \tag{5.21}$$

For the case $P_0 \sim \mathrm{hat}(1, 2)$, $\sigma_l = 0.1$, the individual contributions to $P(k)$ up to second order are plotted in Figure 5-15. The P_{13} contribution is negative ($\sim -k^2$) for $k_1 < k < k_2$ and zero outside this interval; the absolute value of $b^4 P_{13}$ is the dashed curve in Figure 5-15. On large scales, the P_{22} term provides a k^4 tail as $k \to 0$, though this bends over near k_1. Specifically, taking the small k limit of P_{22}, we obtain

$$b^4 P_{22} \to \frac{\pi^3 b^4 k^4}{5k_1^3} = \frac{9\sigma_l^4 k^4}{3920\pi k_1^7} \tag{5.22}$$

This asymptote is plotted as the dot-dashed line in Figure 5-15, and the full $b^4 P_{22}$ term is drawn as the solid curve. For $k_1 < k < k_2$, P_{22} provides a small positive contribution, but $|P_{13}| > P_{22}$, so the linear power is reduced from k_1 to k_2 as power is transferred to smaller scales (and, minimally, to larger scales). From k_2 to $2k_2 = 4k_1$, the power comes from harmonics generated by P_{22}.

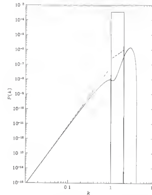


Figure 5-15 Second order contributions to $P(k)$. The linear spectrum (thick solid line) is $b^2P_0 \sim \text{hat}(1,2)$ for $\sigma_l = 0.1$. The smaller nonlinear contributions are b^4P_{22} (solid curve for $0 < k < 4$) and b^4P_{13} (dashed curve for $1 < k < 2$—we plot $|P_{13}|$ since it is negative), and we show the asymptotic k^4 tail (dot-dash line for small k).

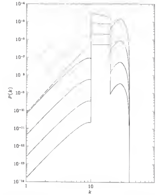


Figure 5-16 Mode coupling and k^4 tail for the second order power spectrum. Based on an initial spectrum $P_0 \sim$ hat(10 20) and perturbation theory up to third order $P(k)$ is plotted for $\sigma_l = 0.125$, 0.25, 0.5, 1.0, 2.0 and 4.0 (bottom to top). The asymptotic k^4 tail (dashed line) is plotted in the $\sigma = 1.0$ case. Perturbation theory breaks down for $\sigma_l > 1$ (dotted curves).

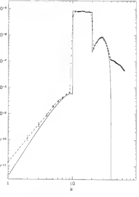


Figure 5.17 Observed mode coupling. Power is measured for initial $P_0 \sim$ hat(10, 20) spectrum evolved to $\sigma_f = 0.5$

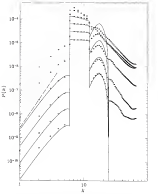


Figure 5-18 Measured power for initial $P_0 \sim$ hat(6, 12). The six epochs are specified by $\sigma_l = (1.5)^i$, $i = -3, \ldots, 2$ (spanning from $\sigma_l \approx 0.3$ to 2.25 from bottom to top). Perturbation theory has broken down for the dotted curves ($\sigma_l > 1$). The asymptotic k^4 tail (dashed line) is plotted for $\sigma_l = 1$.

Recall that k has already been scaled by $2\pi/L$ to account for the integer wavevectors used on the simulation grid. For our simulations, $k_1 > 1$, but k_1 can not be too large, since we must account for the dynamic range: 128^3 grid, $k_{nyq} = 64$, single precision code (real numbers are four bytes). To illustrate the scaling with time, the perturbation theory calculation is shown in Figure 5-16 for the case $k_1 = 10$, with evolution specified by $\sigma_l = 0.125$, 0.25, 0.5, 1.0, 2.0, and 4.0. For $\sigma_l \approx 1.47$, the P_{13} begins to overcome the linear contribution and perturbation theory breaks down. The dotted curves ($\sigma_l = 2$ and 4) in Figure 5-16 are therefore truncated. Measured power in the perturbative regime is shown in Figure 5-17, where $\sigma_l = 0.5$ for an initial hat(10,20) spectrum. The spectrum shown was measured from one realization, rather than averaging over several to reduce noise. This $\sigma_l = 0.5$ run is just early enough that the $2k_1 < k < 4k_1$ range is well covered by the 1-loop mode coupling. Later on, 2-loop effects (P_{24}) decrease $P(k)$ in the $2k_1 < k < 4k_1$ range while providing (P_{33}) leading order harmonics for $4k_1 < k < 6k_1$ (Fry 1994a). For small k the signal is modified by aliasing—spurious high k (above Nyquist) to low k transfer due to the periodicity of the FFT series. To better observe the k^4 tail and the k^2 and k^4 scaling in a simulation, we choose $k_1 = 6$, to allow grid effects to be washed out even for the early ($\sigma_l \lesssim 1$) epochs. The measured power spectrum and the perturbative result are shown in Figure 5-18 for the hat(6,12) case, at the six times $\sigma_l = 0.296$, 0.444, 0.667, 1, 1.5, and 2.25 (which differ a constant factor of 1.5). The signal has been cleaned up by averaging over 20 realizations of the initial Gaussian field. The aliasing of the FFT cannot be eliminated here. The small-scale signal for $k > k_{nyq} = 64$ is folded back into the $1 \leq k \leq 64$ range. This effect is apparent for $k \approx k_{nyq}$ (note that the signal levels off or even turns upward). Also, the aliasing is apparent at small k since the magnitude of the perturbative k^4 part is comparable to or smaller than the measured signal leveling off at the Nyquist frequency (note that for $k < 3$ this is true in all the curves of Figure 5-18).

We performed high resolution one-dimensional simulations to check whether higher order perturbative effects can be detected in the power spectrum. Using a $2^{18} = 262144$ grid size and a hat(10, 20) initial spectrum, the power spectrum was measured by averaging $|\delta(\mathbf{k})|^2$ for 100 realizations. For six times $\sigma_l = 2^i$, $i = -5, -4, \ldots, 0$ (from $\sigma = 0.03125$ to 1), the measured power is plotted in Figure 5-19. The second order power spectrum (the 1-d version), which includes the k^4 tail and harmonics up to $k = 40$, is also shown at each epoch. Note that the second order spectrum (dotted curve) for the $\sigma_l = 1$ case is truncated. Higher order mode coupling is clearly evident. The $P \sim \delta^{2(n-1)}$ scaling for n-loop corrections is illustrated in Figure 5-20, which uses the same data as Figure 5-19 but plotted at several fixed k values to show the scaling with time. We have $k_1 = 10$ and $k_2 = 20$, and for $k > 20$ we choose k values of 25, 45, and 65 to test the 1-, 2-, and 3-loop scaling, since P_{mn} provides the leading contribution from $(m-1)k_2$ to mk_2 (Fry 1994a).

The P_{mn} contributions to $P(k)$ may be integrated to obtain the local variance

$$\sigma^2 = \langle \delta^2 \rangle = \int d^3k \, P(k) \tag{5.23}$$

For the case $P_0 \sim k^n$, sharp cutoff at k_c, Scoccimarro and Frieman (1996) have obtained the higher order corrections up to the 3-loop level (including up to seventh order in the perturbation series (5.19)). Scoccimarro and Frieman (1996) find that, independent of n,

$$\begin{aligned} \sigma^2 &= \sigma_l^2 + \frac{19}{15}\sigma_l^4 + \frac{91}{45}\sigma_l^6 + \frac{1477}{375}\sigma_l^8 + \mathcal{O}(\sigma_l^{10}) \\ &= \sigma_l^2 + 1.267\,\sigma_l^4 + 2.022\,\sigma_l^6 + 3.939\,\sigma_l^8 + \mathcal{O}(\sigma_l^{10}) \end{aligned} \tag{5.24}$$

We find that while $d^2\sigma/d\sigma_l^2 \gtrsim 0$, this expression provides a good approximation to the measured values (provided that normalization integrals are computed using $\sum_k$

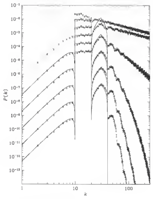


Figure 5-19 Mode coupling for 1-d hat(10 20) simulation Measured power averaged over 100 simulations, is shown at six epochs $\sigma_l = 2^i$ $i = -5$ 0 (from ≈ 0.03 to 1) The second order power spectra plotted as the solid curves For $\sigma_l = 1$ (dotted curve) perturbation theory breaks down Higher order perturbative mode coupling is visible for $k > 40$

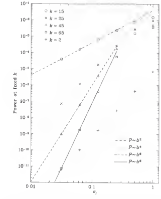


Figure 5-20. Perturbative scaling in the 1-d power spectrum. From the data of Figure 5-19 we show power versus time $\sigma_l \sim b(t)$ at several fixed wavenumbers. The linear order (tree level) spectrum scales according to $P \sim b^2$ while the n loop contributions scale as $b^{2(n+1)}$. The lines of slope 2, 4, 6 and 8 indicate these growth laws.

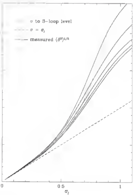


Figure 5.21. Local fluctuations to 3-loop level for $P_0 \sim k^n$. The solid curves show measured rms fluctuations for $n = -3$ (top curve), -2, -1, 0, and 1 (bottom curve) for an initial power law spectrum with $k_c = 10$. The dashed line shows the first order result. The dotted curve shows the perturbative expansion for σ including 3-loop contributions.

[illegible] for the discrete simulation grid). In Figure 5-21 we plot the (5.24) result along with measured values. An initial k^n spectrum with $n = -3, -2, -1, 0, 1$ and $k_c = 10$ was used.

5.2.3 Reduced Bispectrum

We have measured 3-point correlations under a variety of conditions. In particular, the bispectrum $B(k_1, k_2, k_3)$ has been measured for underlying matter distributions and discrete samples. The three degrees of freedom associated with the closed k space triangle need to be conveniently parametrized. One natural choice (Fry, Melott, & Shandarin 1993) is to fix magnitude of the $\mathbf{k}_1$ and $\mathbf{k}_2$ vectors, where we take $k_1 > k_2$. Then, since $\sum_i \mathbf{k}_i = 0$, the third vector has a magnitude which spans the range $k_1 - k_2$ to $k_1 + k_2$, and we parametrize the configuration by θ, the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$. The $k_1 \pm k_2$ range in length scales allows us to gauge nonlinearity by setting k_1 with respect to k_{nl}, while the θ dependence reflects the anisotropic nature of gravitational instability.

Second order perturbation theory provides a useful prediction for the bispectrum using the G_2 kernel in (5.18)

$$
\begin{aligned}
B(k_1, k_2, k_3) &= Q_{12} P_1 P_2 + Q_{23} P_2 P_3 + Q_{31} P_3 P_1 \\
Q_{ij} &= 1 + \mathbf{k}_i \cdot \mathbf{k}_j \left(\frac{k_i}{k_j} + \frac{k_j}{k_i} \right) + (\mathbf{k}_i \cdot \mathbf{k}_j)^2
\end{aligned}
\tag{5.25}
$$

This is the perturbative result to first nonvanishing order, i.e. to tree level (Fry 1984). A systematic development of 3-point correlation results, including 1 loop corrections, is provided by Scoccimarro (1997).

Inspired by galaxy survey measurements (Peebles & Groth 1975, Fry & Seldner 1982) which show hierarchical 3-point correlations and by theory, the reduced bispectrum Q is defined as

$$
Q(k_1, k_2, k_3) = \frac{B(k_1, k_2, k_3)}{P(k_1)\,P(k_2) + P(k_2)\,P(k_3) + P(k_3)\,P(k_1)} \tag{5.26}
$$

With this definition, much of the B dependence on scale, time, and initial spectrum is eliminated, leaving a function with values of order unity. For example, with equilateral triangles ($\mathbf{k}_i \cdot \mathbf{k}_j = -\frac{1}{2}$) the reduced bispectrum (5.26) is a constant $\frac{4}{7}$. For the reasons mentioned above, we utilize anisotropic triangle configurations parametrized by the angle θ between $\mathbf{k}_1$ and $\mathbf{k}_2$. We fix $k_1 = 2k_2$, so the range in k goes from $\frac{1}{2}k_1$ to $\frac{3}{2}k_1$. For scale-free initial spectra, $P_0 \sim k^n$ with no cutoff (or with $\frac{3}{2}k_1$ below the filtering scale), equation (5.26) gives a bispectrum with a significant shape dependence, as shown in Figure 5-22.

Although we specify $k_2/k_1 = 0.5$, it is possible to choose other values for the bispectrum triangle configuration. At the least, a clean $P(k)$ signal at k_1, k_2 and for the $k_1 - k_2 < k_3 < k_1 + k_2$ range would be needed. If the 2-point function is not clear, one would not expect the 3-point function to yield much that is useful. In terms of the dynamic range of our simulations, $k_2/k_1 = 0.5$ provides a good compromise. If one chooses $k_2/k_1 \ll 0.5$, then essentially two scales are being probed, $k_1 \approx k_3$ and large scales k_2, both of which are likely not to be within range of measurement in the same data set. Choosing $k_2 \approx k_1$ gives triangles which span scales $0 < k < 2k_1$, a viable possibility only if signals are clear in the entire range. In Figure 5-23 we show the $Q(\theta)$ perturbative predictions for various k_2/k_1 ratios, for $P_0 \sim k^{-1}$.

From a signal detection point of view, the more shape dependence the better. In the nonlinear regime, we find that the $Q(\theta)$ curve tends to flatten out. We observe a similar flattening when bias is present. In trying to ascertain the effects on nonlinearity or bias, we use the tree level perturbative result $Q_{pt}(\theta)$, equation (5.25), as our reference. As for obtaining more or less configuration dependence, it turns out that the k_2/k_1 choice is not necessarily crucial, which can be seen from Figure 5-23, though we find $k_1 = 2k_2$ the most practical choice. Although a scale-free k^n spectrum is an idealization of sorts, it is quite possible to find data for which effectively

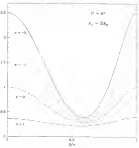


Figure 5 22 Reduced bispectrum $Q(\theta)$ in perturbation theory for scale-free power $P \sim k^n$ with no cutoff The Fourier space triangles are defined by $k_1 = 2k_2$ and θ is the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$ From top to bottom the curves are for $n = -2$ -1 0 and 1

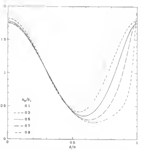


Figure 5.23 Perturbative $Q(\theta)$ for various k_2/k_1 values For triangle configurations specified by a fixed k_2/k_1 ratio (θ is the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$) the reduced bispectrum is plotted for $P \sim k^{-1}$

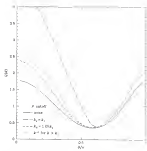


Figure 5-24. Perturbative $Q(\theta)$ with $k_1 = 2k_2$ and power filter $k_f \approx k_1$. The power spectrum is $P \sim k^{-1}$ up to the filtering scale, after which it is cut off or approximated by k^{-3}

$P \sim k^n$ over the length scales studied. If $k_1 \le k_f$ but $\frac{3}{2}k_1$ falls beyond k_f, then equation (5.25) could still be applied to get nonzero Q predictions, but this is not useful. The measured power can be used in the $Q(\theta)$ equation (5.25), or one can use an approximate model. A useful estimate takes the linear power for $k < k_f$ and joins this with an approximate $P \sim k^{-3}$ beyond, since we found that our k_{nl} filtered simulations typically had such a high k tail. In Figure 5.24 we show the bispectrum curve for $P \sim k^{-1}$ with various high k behaviors when $k_f \approx k_1$. Along with the scale-free case, we show a sharp cutoff at k_1, a Gaussian cutoff at $1.25k_1$, and a sharp transition to k^{-3} at k_1.

In an effort to have a strong signal, we let k_1 coincide with k_{nl}, which also serves as the filtering scale. It is possible, and even desirable, to choose $k_1 < k_{nl}$. But if k_1 is too small, two difficulties may arise. The bispectrum signal may become indistinguishable from the noise, since the bispectrum is small in the linear regime. Also, for small k_1 values, there are less triangle configurations in k-space to average over. Fry et al. (1993) estimate this statistical effect on the reduced bispectrum error bars ΔQ in three dimensions to roughly follow $\Delta Q \sim (k_{nl}/k_1)^{3+n}$, qualitatively born out in our data. We choose four epochs, $k_{nl} = 32$, 16, 8, and 4, in a series of $P_0 \sim k^n$ simulations with $n = 1$, 0, -1, and -2. Results are shown in Figures 5-25 through 5-28 in the case of sharp $k_c = k_{nl}$ filtering. Analogous Gaussian filtering results are shown in Figures 5-29 through 5-32. The Q error bars shown are the statistical errors associated with averaging over k-space configurations. Even at this stage of nonlinearity, the Q signal shows the characteristic shape expected from perturbation theory, and furthermore the curves are more flattened as n increases. For completeness, we show the data for each of the four different cutoffs. Were it not for the k-space resolution differences, one would expect the signals to be virtually identical (for the same n)

since they are characterized by $k_{nl} = k_1$ in each case. Figure 5-33 shows the bispectrum data for the four different cutoffs in the $P_0 \sim k^{-1}$ case. The perturbative $Q(\theta)$ curve is plotted using $P \sim k^{-3}$ for $k > k_c$. Perhaps surprisingly, this accounts for the overall behavior of the data, though the measured curve is more flattened.

The clustering signal in the reduced bispectrum, which we have called $Q(\theta)$, is obtained from the raw signal by performing discreteness and sharpening corrections. The raw bispectrum is $\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle$ averaged over closed triangle $(\sum \mathbf{k}_i = 0)$ configurations. Taking into account shot noise, this is related to the clustering 3-point signal $B(k_1, k_2, k_3)$ by

$$\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle = B(k_1, k_2, k_3) + \frac{1}{V}[P(k_1) + P(k_2) + P(k_3)] + \frac{1}{V^2} \tag{5.27}$$

This situation is analogous to the 2-point case, equation (5.3). Let $\mathbf{k}_1$ and $\mathbf{k}_2$ be nonzero, with $k_1 > k_2$. Then to close the triangle we have

$$k_3^2 = k_1^2 + k_2^2 + 2k_1k_2\cos\theta \tag{5.28}$$

The nonzero restrictions simply remove $\mathbf{k}_i = 0$ delta function terms (not applicable to the measurements anyway) from the expression below. Directly from the discrete grid counts array $m(\mathbf{x})$, we calculate the third moment to be

$$\begin{aligned}\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle &= \frac{1}{V^2}\sum_{\mathbf{r},\mathbf{r}'}\zeta(\mathbf{r},\mathbf{r}')e^{-i(\mathbf{k}_1\cdot\mathbf{r}+\mathbf{k}_2\cdot\mathbf{r}')} \\ &+ \frac{\langle m^2\rangle V}{N^2}\frac{1}{V}\sum_{\mathbf{r}}\xi(r)[e^{-i\mathbf{k}_1\cdot\mathbf{r}} + e^{-i\mathbf{k}_2\cdot\mathbf{r}} + e^{-i\mathbf{k}_3\cdot\mathbf{r}}] + \frac{\langle m^3\rangle V}{N^3}\end{aligned} \tag{5.29}$$

where ζ is the spatial 3-point function. The two discreteness terms involving $1/N$ (or $1/N_{\mathrm{eff}}$) reproduce the $1/V$ terms of (5.27) in the appropriate limits. As in the power spectrum case, we find that discreteness corrections (5.27) are reasonably effective in separating out the clustering signal. Isotropic sharpening corrections are applied after the shot noise subtractions.

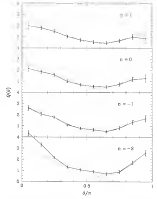


Figure 5-25 Measured reduced bispectrum $Q(\theta)$ for $P_0 \sim k^n$ with sharp cutoff $k_c = k_{nl} = 32$ and triangles with $k_1 = 2k_2 = 32$. The perturbative results Q_{pt} using the measured power are shown as the dotted curves.

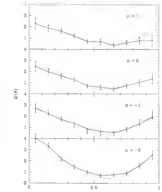


Figure 5-26 Measured reduced bispectrum $Q(\theta)$ for $P_0 \sim k^n$ $k_c = k_{nl} = 16$ and $k_1 = 2k_2 = 16$ Dotted curves Q_{pt} using the measured power

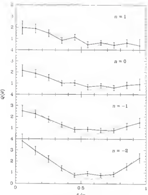


Figure 5-27 Measured reduced bispectrum $Q(\theta)$ for $P_0 \sim k^n$ $k_c = k_{nl} = 8$ and $k_1 = 2k_2 = 8$ Dotted curves Q_{pt} using the measured power

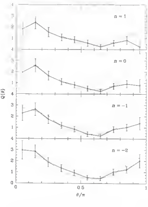


Figure 5-28 Measured reduced bispectrum $Q(\theta)$ for $P_0 \sim k^n$ $k_c = k_{nl} = 4$ and $k_1 = 2k_2 = 4$ Dotted curves Q_{pt} using the measured power

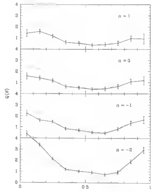


Figure 5-29 Measured reduced bispectrum $Q(\theta)$ for $P_0 \sim k^n$ with Gaussian cutoff $k_G \approx k_{nl} = 32$ and triangles with $k_1 = 2k_2 = 32$. The perturbative results Q_{pt} using the measured power are shown as the dotted curves.

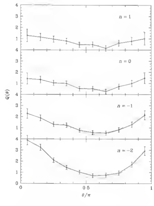


Figure 5-30 Measured reduced bispectrum $Q(\theta)$ for $P_0 \sim k^n$ $k_G \approx k_{nl} = 16$ and $k_1 = 2k_2 = 16$ Dotted curves Q_{pt} using the measured power

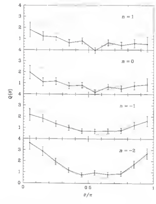


Figure 5-31 Measured reduced bispectrum $Q(\theta)$ for $P_0 \sim k^n$ $k_G \approx k_{nl} = 8$ and $k_1 = 2k_2 = 8$ Dotted curves Q_{pt} using the measured power

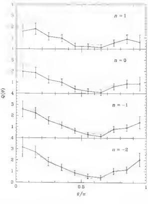


Figure 5-32 Measured reduced bispectrum $Q(\theta)$ for $P_0 \sim k^n$ $k_G \approx k_{nl} = 4$ and $k_1 = 2k_2 = 4$ Dotted curves Q_{pt} using the measured power

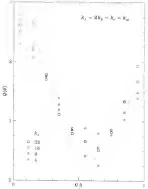


Figure 5-33 Bispectrum $Q(\theta)$ for various $k_c = k_1$ power spectrum cutoffs $P_0 \sim k^{-1}$ up to k_c. The curve shows the tree-level perturbative result for $P \sim k^{-1}$ for $k < k_c$ $\sim k^{-3}$ for $k > k_c$

The measured power generally follows the linear theory predictions for $k < k_{nl}$. Results from §5.2.3 suggest that tree level perturbation theory is still providing qualitatively correct bispectra predictions into the nonlinear regime. Linear theory predicts $Q = 0$, but the skewness of the δ distribution increases with time ($\sim b(t)$ to lowest order), and we find that a 3-point signal is measurable even in the quasi-linear regime. For the 3-point function, second order perturbation theory is utilized to find the first nonvanishing predictions. We may control the nonlinearity in our $Q(\theta)$ measurements by varying k_1/k_{nl}. Choosing $k_1 \ll k_{nl}$, though not possible in practice due to limited dynamic range, would guarantee compliance with perturbation theory. Since $k_1 = 2k_2$, choosing $\frac{3}{2}k_1 \lesssim k_{nl}$ should perhaps prove sufficient for some second order perturbative results to play a useful role in data analysis involving mild nonlinearities and bias.

Since the initial spectra are filtered for $k \gtrsim k_f$, the scale-free perturbative results for $Q(\theta)$ are modified if $\frac{3}{2}k_1 \gtrsim k_{nl}$, and we may account for this (e.g. by adding a k^{-3} tail to $P(k)$ or by simply using the measured power in the perturbative Q, as is necessary in astronomical data). Although the tree level perturbative bispectrum expression is fixed for $P_0 \sim k^n$ and $\frac{3}{2}k_1 \lesssim k_{nl}$, the $Q(\theta)$ signal would be expected to change due to higher order nonlinear effects and due to the dynamic range of the Q estimator. We are able to observe higher order effects which depend on the power index n and on k_1/k_{nl}.

The correlation functions should also be modified by biasing. We use a local model of biasing

$$\delta_g = \sum_{n=0}^{\infty} \frac{1}{n!} b_n \delta^n \qquad (5.30)$$

The power spectrum $\langle|\delta_g(\mathbf{k})|^2\rangle$ differs from the underlying $\langle|\delta(\mathbf{k})|^2\rangle$ according to linear bias

$$P_b(k) = b_1^2 P(k) \tag{5.31}$$

If linear bias applied in general, then the 3-point function would be expected to change by the same b_1 factor, $B_b = b_1^3 B$, so the reduced bispectrum would be

$$Q_b = \frac{1}{b_1} Q \tag{5.32}$$

Consistent to second order in perturbation theory and nonlinear local bias, the 3-point function (Fry & Gaztanaga 1993) is

$$Q_b = \frac{1}{b_1} Q + \frac{b_2}{b_1^2} \tag{5.33}$$

We are able to observe the effects of biasing in $P(k)$. More importantly, since there are observational implications (Fry 1994b), we also detect the effects in $Q(\theta)$. While nonlinearities unrelated to bias do affect the $Q(\theta)$ signal, we find that the biasing effects are still appreciable.

Regarding potential observations, note that $Q(\theta)$ is largely Ω_0 independent, due to the nearly exact lack of Ω_0 dependence in the second order perturbative expressions (Bouchet et al. 1992), and it depends on shape rather than normalization. The Hubble constant h and density parameter Ω_0 (which come into play for $P(k)$ normalization) are not directly in the equations. As for the power, note that cosmic background radiation $\Delta T/T$ fluctuations shed light on the largest scales, where $P(k) \sim k$. If uncertainties in the background fluctuation data are low, then (within model dependent uncertainties) the data are really probing ancient, unbiased matter fluctuations. With a large scale $P(k)$ in hand (up to a normalization including an $h^2\Omega_0$ factor), one may try to match this with the galaxy survey power spectra on scales below $100\,h^{-1}$ Mpc, but any determination of b_2 is always tied in with Ω_0 and may be quite model dependent. As pointed out by Fry (1994b) regarding survey data and by Fry et al. (1995) regarding N-body simulation data, the $Q(\theta)$ statistic is a promising candidate for probing bias.

In Figures 5-34–5-38 we control n and k_1/k_{nl} in a series of 3-point measurements. The persistent success of the tree level $Q(\theta)$ result, equation (5.25), is demonstrated. We find the bispectrum signal is flattened by higher order effects for $n \gtrsim -1.5$. For $n \lesssim -1.5$ the configuration dependence is actually more pronounced. The robustness of the perturbative result is not closely tied with use of scale-free power spectra, as we have utilized other initial spectra as well. Furthermore, the filtering we must necessarily perform on the k^n spectra destroys the scale invariance anyway. Each plot shows a dotted curve, calculated from (5.26) using the measured power for that simulation run. For reference, the $k_1/k_{nl} = 0.625$ case also shows the prediction using the k_G filtered P_0 as the dashed curve. These runs use $P_0 \sim k^n$, $k_G \approx k_{nl} = 32$, and we take measurements from one realization. For $k_1/k_{nl} \lesssim 0.5$, the Q signal generally becomes too noisy to be useful. This is illustrated in the $n = 0$ and $n = 1$ cases, Figures 5-37 and 5-38. After evolving the simulations to $k_{nl} = 16$, we show in Figure 5-39 the $k_1/k_{nl} = 0.625$ results, for a sharp $k_c = 16$. The results are similar to those above, though we averaged over 10 realizations to achieve a clean signal.

The question naturally arises as to how biasing in the data set will affect the correlations. Additionally, it is important to know whether nonlinearities unrelated to bias will be so entangled as to make it difficult or impractical to discern the bias unambiguously. We find that the biased power obeys the expected trend: more bias leads to greater measured power. When the power signal is quite clear, we generally find that there are regions (typically in the linear regime) where the biased power is well fit by the linear $b_1^2 P$ relation. At times, a good fit for a constant P_b/P does not exist, either because the signal is too noisy or simply the P_b and P appear

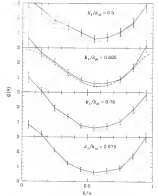


Figure 5-34 Measured reduced bispectrum at various scales for $P_0 \sim k^{-2}$ $k_G = k_{nl} = 32$ Dotted curves: Q_{pt} using the measure power Dashed curve Q_{pt} using the filtered P_0

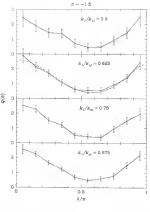


Figure 5.35 Measured reduced bispectrum at various scales for $P_0 \sim k^{-1.5}$, $k_G \approx k_{nl} = 32$. Dotted curves: Q_{pt} using the measure power. Dashed curve: Q_{pt} using the filtered P_0.

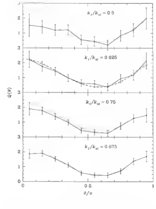


Figure 5-36 Measured reduced bispectrum at various scales for $P_0 \sim k^{-1}$ $k_G \approx k_{nl} = 32$ Dotted curves Q_{pt} using the measure power Dashed curve Q_{pt} using the filtered P_0

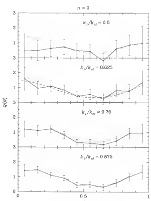


Figure 5.37 Measured reduced bispectrum at various scales for $P_0 \sim k^0$, $k_G \approx k_{nl} = 32$. Dotted curve: Q_{pt} using the measure power. Dashed curve: Q_{pt} using the filtered P_0

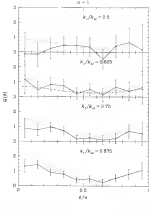


Figure 5-38 Measured reduced bispectrum at various scales for $P_0 \sim k^1$ $k_G \approx k_{nl} = 32$ Dotted curves Q_{pt} using the measure power Dashed curve Q_{pt} using the filtered P_0

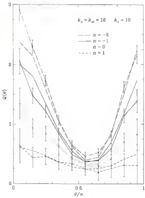


Figure 5-39 Reduced bispectrum for $k_1/k_{nl} = 0.625$ and sharply cut off initial power, averaged over 10 realizations. The Q_{pt} curves, using measured power, are also shown

to have different slopes. We find the bispectrum signal to flatten out with more biasing. Typically, the flattening/shifting according to the nonlinear biasing effect, equation (5.33), provides a useful biasing indicator into the nonlinear regime. We present data in support of these claims below.

Two types of threshold biasing are applied. For sharp threshold bias, cells with $\delta < \delta_t$ are emptied, leaving only the higher density regions (so δ is clipped to -1 for density contrasts below δ_t). For power law bias, the "galaxy" probability function is $\mathcal{P} \sim (1+\delta)^{\alpha}$ below a threshold δ_t. The latter prescription is slightly less drastic, and perhaps more realistic, than the sharp threshold case. We include both for comparison purposes. Especially with the sharp threshold we can achieve a significant bias, such as $\delta_t = 5$ or 10, even within the limited dynamic range of the simulations. The signal degradation is such that we generally use $\delta_t \lesssim 6$.

Relatively clean biased power signals are shown in Figure 5-40, obtained from 10 realizations of $P_0 \sim k^n$, $k_c = 16$ simulations. For each initial spectrum, sharp and $\rho^{1/2}$ thresholds are shown, for $\delta_t = 3$ and 6. Fitting to the $b_1^2 P$ form for $k < k_{nl}$, we show the linear bias parameter b_1 in Figure 5-41 as a function of δ_t. The Jensen and Szalay (1986) result for clipped Gaussian fields, equation 4.39, is shown as the dotted line (we use $\sigma_t = 1$ in the denominator). The functional relationship is steeper in the Gaussian case, as might be expected since clustered data certainly have a $\sigma > 1$ non-Gaussian tail. There is not much δ_t dependence in the $\mathcal{P} \sim \rho^{1/2}$ case, since for $\delta_t \gtrsim 1$ one might expect more dependence on the exponent α. There is just enough evolution at $k_{nl} = 32$ in our simulations to use $k_G \approx 32$ filtered spectra and utilize the $k < k_{nl}$ signal. Figure 5-42 shows the biased power for one realization, again for two power law and two sharp threshold biases (note the high k problems). Choosing $k_{nl} = 16$ generally allows enough evolution for utilizing several biases while still maintaining a significant small k linear regime. Particularly clean biasing signals are shown in

Figure 5-43, which shows the power for 20 realizations of a $P_0 \sim k^{-1.5}$, $k_c = 16$ simulation. With a ρ^{α} threshold, $\delta_t = 4$, we show the unbiased power (the lowest solid curve) as well as $\alpha = 0.1$, 0.5, 1, and 2 biased samples (the dotted curves). Additionally, a sharp $\delta_t = 10$ sample is shown (dashed curve). For the $\delta_t = 10$ case, we also show the signal from just one realization (the sample had only $N = 6000$ galaxies). As one might expect, in the mildly nonlinear regime, averaging over many realizations (analogous to using very large astronomical data sets) really does narrow in on the signal. We plot the linear bias parameter b_1 versus the threshold power index α in Figure 5-44.

We now turn to results concerning nonlinear bias and the bispectrum. In general the results are quite encouraging in the mildly nonlinear regime, as the effects of biasing are detectable using the results of second order perturbation theory.

Using initial power indices of $n = -2$, -1, 0, and 1, we demonstrate nonlinear biasing effects in Figures 5-45 through 5-52. We use $k_{nl} = k_c = 16$ simulations and $k_1 = 2k_2 = 10$, so the bispectrum $Q(\theta)$ will be probing the quasilinear regime. Measurements were obtained from one realization of the initial conditions. Figure 5-45, for $n = -2$, shows the unbiased reduced bispectrum (the solid curve connecting the data points) and two bispectra for two different biasing schemes. The data points from the $\mathcal{P} \sim \rho^{1/2}$, $\delta < 3$ prescription are well fit by the dotted curve, the least squares fit of the measured Q_b to the measured Q using the nonlinear biasing form $Q/b_1 + b_2/b_1^2$. Similarly, the data points for the sharp $\delta_t = 3$ prescription are closely followed by the short-dashed curve. Additionally, we plot (long-dashed curve) the same nonlinear bias fit using the Q_{pt} generated by $P \sim k^{-2}$ rather than the measured Q. This emphasizes that the nonlinear biasing effect is larger than the nonlinear effect which slightly increases Q variation (over the tree-level Q_{pt} θ dependence) for $n = -2$. Using the measured Q in fitting to the measured Q_b is utilizing more information than

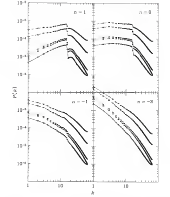


Figure 5-40 Biased power spectrum measurements. For initial k^n, $k_c = 16$ spectra, we show the unbiased power and the power from four biased samples, averaged over 10 realizations. In each panel, from bottom to top we have: unbiased (solid curve), $\mathcal{P} \sim \rho^{1/2}$ up to $\delta_t = 3$ and 6 (dotted curves), and sharp $\delta_t = 3$ and 6 thresholds (dashed curves).

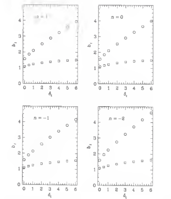


Figure 5-41 Linear bias parameter b_1 from power spectrum measurements. For initial k^n $k_c = 16$ spectra, we plot b_1 versus the bias threshold δ_t. The circles apply to a sharp threshold, the squares to a $\mathcal{P} \sim \rho^{1/2}$ scheme below δ_t. The dotted line shown in the $n = -1$ case is the analytic result for the clipping of a Gaussian field

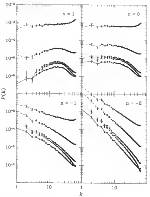


Figure 5-42 Biased power spectrum measurements. For initial k^n, $k_G \approx 32$ spectra we show the unbiased power and the power from four biased samples averaged over 1 realization. In each panel, from bottom to top we have: unbiased (solid curve), $\mathcal{P} \sim \rho^{1/2}$ up to $\delta_t = 2$ and 6 (dotted curves) and sharp $\delta_t = 2$ and 6 thresholds (dashed curves).

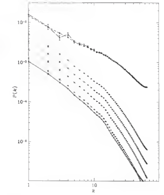


Figure 5.43 Biased power spectrum for initial $k^{-1.5}$, $k_c = 16$ spectrum, averaged over 20 realizations. The lowest (solid) curve is the unbiased spectrum. The dotted curves correspond to power law ($\mathcal{P} \sim \rho^{\alpha}$) thresholds with $\delta_t = 4$. From bottom to top $\alpha = 0.1$, 0.5, 1 and 2. The short dashed curve the sharp clipping $\delta_t = 10$ signal while the long-dashed line near it comes from one dilute realization.

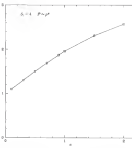


Figure 5-44. Linear bias parameter b_1 from power measurements based on 20 realizations of initial $k^{-1.5}$ spectra. For a power law threshold $\mathcal{P} \sim \rho^{\alpha}$, below $\delta_t = 4$ the index α is varied

would be available in a galaxy survey. Even in ideal observational data, one would only have a measured P and a measured Q. Given these spectra, one would ask whether they were consistent with no biasing or whether they were really P_b and Q_b with measurable bias parameters. We utilize the full information, as that is one the main benefits of a computer simulation approach (we do intentionally limit our information below, when we demonstrate the viability of the bispectrum technique in a dilute sample). To illustrate the biasing detection for various additional prescriptions we plot the Q/b_1+c fitting parameters in Figure 5-46 as a function of the δ_t threshold. The b_1 biasing parameter from Q_b is generally comparable to that obtained from P_b, though typically we find that $b_1 > b_P$ when $n \gtrsim -1.5$, where we use b_P to denote the $P_b = b_P^2 P$ fitting parameter. Furthermore, in the mildly nonlinear regime we find that while $b_1 \not\approx 1$ (accounting for the 1-σ error bars), the nonlinear biasing form $Q/b_1 + b_2/b_1^2$ is consistently a much better fit that the naive Q/b_1. For example, in the sharp $\delta_t = 3$ biasing for $n = -2$, the power spectra gave $b_P = 3.25 \pm 0.01$ with χ^2 per degree of freedom $\chi^2/15 = 1.23$. The strict linear $P_b = b_P^2 P$ fits for $k < k_{nl}$ are generally not good by this criterion. If one simply uses the ratio of P_b to P over the modes sampled by the bispectrum, then the discrepancy between b_P and b_1 still exists but is often smaller. The nonlinear biasing fit gives $b_1 = 2.9 \pm 0.14$, $b_2/b_1^2 = 0.24 \pm 0.02$ with $\chi^2/8 = 0.34$, while the Q/b_1 fit gives $b_1 = 2.00 \pm 0.03$ with $\chi^2/8 = 1.7$. While in the mildly nonlinear regime, the Q/b_1 fit worsens relative to the $Q/b_1 + b_2/b_1^2$ fit as the biasing is increased. Similar considerations hold for the $n = -1$ case, Figures 5-47 and 5-48. For $n = 0$ and $n = 1$, Figures 5-49 through 5-52, the signal has degraded yet bias is detected.

We demonstrate the utility of large data sets in Figure 5-53, where we average over 20 realizations of an initial $k^{-1.5}$, $k_{nl} = k_c = 16$ spectrum. Two cleanly fit biased bispectra are shown, for $\mathcal{P} \sim \rho^{1/2}$ prescription (dotted curve) and for a $\delta_t = 10$

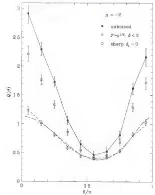


Figure 5-45 Biased bispectrum for initial power index $n = -2$ The unbiased signal (solid curve) and the Q_3 for two biasing prescriptions with threshold $\delta_t = 3$ are shown Open squares bias probability $\mathcal{P} \sim \rho^{1/2}$ Open circles sharp threshold Dotted and dashed curves best fits to nonlinear bias form

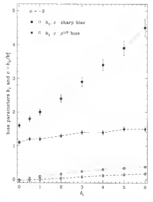


Figure 5.46 Biasing parameters b_1 and $c = b_2/b_1^2$ for power index $n = -2$ as a function of threshold δ_t.

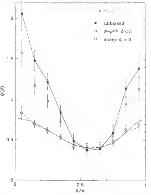


Figure 5-47 Biased bispectrum for initial power index $n = -1$ The unbiased signal (solid curve) and the Q_b for two biasing prescriptions with threshold $\delta_t = 3$ are shown Open squares bias probability $\mathcal{P} \sim \rho^{1/2}$ Open circles sharp threshold Dotted and dashed curves best fits to nonlinear bias form

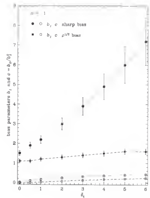


Figure 5-48 Biasing parameters b_1 and $c = b_2/b_1^2$ for power index $n = -1$ as a function of threshold δ_t

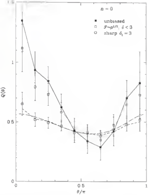


Figure 5-49 Biased bispectrum for initial power index $n = 0$. The unbiased signal (solid curve) and the Q_b for two biasing prescriptions with threshold $\delta_t = 3$ are shown. Open squares: bias probability $\mathcal{P} \sim \rho^{1/2}$. Open circles: sharp threshold. Dotted and dashed curves: best fits to nonlinear bias form.

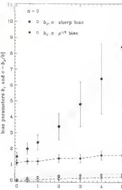


Figure 5-50 Biasing parameters b_1 and $c = b_2/b_1^2$ for power index $n = 0$ as a function of threshold δ_t

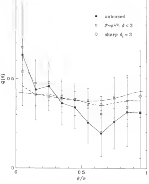


Figure 5-51 Biased bispectrum for initial power index $n = 1$ The unbiased signal (solid curve) and the Q_b for two biasing prescriptions with threshold $\delta_t = 3$ are shown Open squares bias probability $\mathcal{P} \sim \rho^{1/2}$ Open circles sharp threshold Dotted and dashed curves best fit to nonlinear bias form.

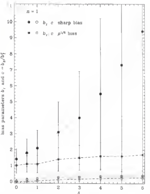


Figure 5-46 Biasing parameters b_1 and $c = b_2/b_1^2$ for power index $n = 1$ as a function of threshold δ_t

(dashed curve). The rather large threshold bias demonstrates how one may find a flat bispectrum in the quasi-linear regime. The nonlinear biasing fit gives $b_1 = 1.0 \pm 1.1$, $b_2/b_1^2 = 0.52 \pm 0.01$ ($\chi^2/\nu = 0.41$). Figure 5-54 shows the sensitivity of Q_b to the biasing prescription, as we set $\alpha = 0.1, 0.3, 0.5, 0.7, 0.9, 1.0, 1.5$, and 2.0 for a $\mathcal{P} \sim \rho^{\alpha}$ bias scheme, with biasing parameters displayed in Figure 5-55. This degree of sensitivity is potentially useful in astronomical data sets, to the extent that there was certainty in that one was working in the quasi-linear regime.

If we let $k_1 > k_{nl}$, ostensibly probing the nonlinear regime with the bispectrum, then a well defined 3-point signal is certainly expected since the matter distribution is far from Gaussian. For k_1/k_{nl} values of 2, 3, 4, etc. in our simulations, we are really probing the shell crossing or catastrophe/singularity sector, much as we were in observing the approximate k^{-3} high k tail to the power spectrum. As such, this area does not really mimic gravitational instability. The pancakes (and filaments) would lead to the now familiar preference (enhanced correlation) for collinear triangle configurations in k-space. We show this in Figure 5-56, a representative example with $k_1 = 24 = 3k_{nl}$. Note that there is a range where Q is flat (similar, coincidently, to the flatness one expects from astronomical observations and N-body simulations in the very highly nonlinear regime). The measured signal (solid curve in Figure 5-56) is not closely tied with the perturbative result (dashed curve, obtained from the measured power). The biased bispectrum, however, is still well fit to the $Q/b_1 + c$ form (dotted curve), giving $b_1 = 4.2 \pm 0.2$, $c = 0.13 \pm 0.03$.

We conclude this chapter with results pertaining to dilute samples. Using only the measured power spectrum and bispectrum of a sparse sample, we demonstrate the degradation of signal with decreasing N (the number of galaxies) and the detection of bias. In Figures 5-57 and 5-58 we show power and bispectrum measurements utilizing only the information from $N \approx 10^4$ and $N \approx 10^3$ data sets. We use $P_0 \sim k^{-1.5}$

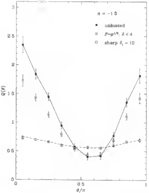


Figure 5-53 Biased bispectrum for initial power index $n = -1\ 5$ The unbiased signal (solid curve) and the Q_b for two biasing prescriptions are shown averaged over 20 realizations Open squares, bias probability $\mathcal{P} \sim \rho^{1/2}$ $\delta < \delta_t$ Open circles sharp threshold $\delta_t = 10$ Dotted and dashed curves best fits to nonlinear bias form.

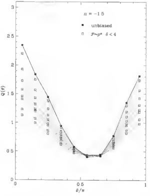


Figure 5-54 Biased bispectrum for initial power index $n = -1.5$. The unbiased signal (solid curve) and the Q_5 for eight biasing prescriptions, $\mathcal{P} \sim \rho^{\alpha}$ $\delta < \delta_t = 4$ are shown (from 20 realizations) The α values range from 0.1 (top dotted curve) to 2.0 (bottom dotted curve)

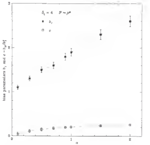


Figure 5-55 Biasing parameters b_1 and $c = b_2/b_1^2$ for power index $n = -2$ as a function of index α for $\mathcal{P} \sim \rho^\alpha$, $\delta < \delta_t$.

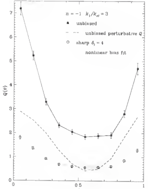


Figure 5-56 Bispectrum in nonlinear regime $k_1 = 3k_{nl}$ Solid curve: measured unbiased Q Dashed curve: perturbative unbiased Q Dotted curve: best fit to nonlinear bias form using measured Q

$k_{nl} = k_c = 16$ simulations $k_1 = 10$ and a sharp $\delta_t = 4$ bias threshold. The full set of information (averaged over 20 realizations) is displayed in Figure 5-59 with the fitting parameters shown in Figure 5-60 as a function of N. The biased bispectrum signal remains robust for $N \gtrsim 4 \times 10^3$ as shown in Figure 5-61, for a mildly biased sample probed at $k \lesssim k_{nl}$

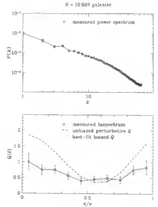


Figure 5-37 Detecting bias for $N \approx 10^4$ galaxies. Using only the measured (biased) power spectrum and bispectrum, the nonlinear bias parameters $b_1 = 3.1 \pm 0.9$ and $b_2/b_1^2 = 0.30 \pm 0.08$ are detected.

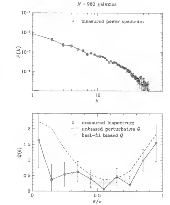


Figure 5-58 Detecting bias for $N \simeq 10^3$ galaxies Using only the measured (biased) power spectrum and bispectrum the nonlinear bias parameters $b_1 = 2.1 \pm 1.0$ and $b_2/b_1^2 = 0.1 \pm 0.2$ are detected

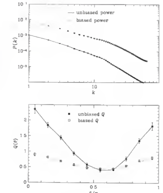


Figure 5-59 Detecting bias under ideal conditions The initial spectrum is $P_0 \sim k^{-1.5}$, $k_{nl} = 16$ as in Figures 5-57 and 5-58 though the data have been averaged over 20 realizations (each containing $N \approx 10^5$ galaxies) Fitting the biased bispectrum gives $b_1 = 4.1 \pm 0.2$ and $b_2/b_1^2 = 0.36 \pm 0.02$ using the measured Q and Q_b

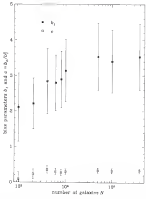


Figure 5-60 Biasing parameters for sparse samples The nonlinear bias parameters b_1 and b_2/b_1^2 are measured as the number of galaxies N is reduced

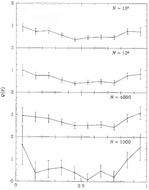


Figure 5-81 Biased bispectrum as the number of galaxies N is reduced

CHAPTER 6
FINAL REMARKS

Throughout this study, our goal has been to illustrate the nature of large-scale structure in the universe by utilizing a perturbative approach to gravitational instability. The applicability of a relatively simple expanding fluid model is somewhat of a luxury, as the nonlinear nature of the large-scale structure problem could conceivably have limited perturbation theory to a very minor role (though the smallness of the measured $\Delta T/T \approx 10^{-5}$ microwave background fluctuations provides a solid foundation for such an approach). Perturbation theory is not the whole answer, but on large scales it may allow us to understand our part of the universe and how—or if—this neighborhood fits within some as yet undetermined robust cosmological model.

The standard model of cosmology is a broad theory which utilizes various branches of physics and which pertains to potential measurements on many scales, from globular clusters and galactic halos to "fair samples" of the universe containing billions of galaxies. Within cosmology, we have studied large-scale structure, a field which generally involves galaxy formation, gravitational instability, the nature of dark and luminous cosmic matter, and the description of the large-scale matter distribution. These ingredients are interrelated. We have restricted our study to regions which are large enough that gravity plays the dominant role but which fall within the subhorizon, slow motion limit of general relativity. This is not a severe restriction, in that a Newtonian approximation may apply from galaxy-sized regions to billion light-year regions which may comprise fair statistical samples of the universe. We applied the expanding fluid model in the linear and mildly nonlinear regimes to attack a very complex problem in a controlled fashion. Our approach, which utilized the efficient

Lagrangian mappings introduced by Zel'dovich (1970), is complementary to important methods involving full N body simulations. The perturbative approach to the equations of motion has been expounded by Lifshitz (1946), Zel'dovich (1970), Peebles (1980), Fry (1984), Goroff et al. (1986), Bertschinger (1992), and others. The Zel'dovich approximation (Zel'dovich 1970, Shandarin & Zel'dovich 1989) is not the same as the exact gravitational instability theory, though it has its own perturbation theory (Grinstein & Wise 1987), similar in structure to the exact perturbative expansion (but with simpler kernels), and this has allowed us to probe the development of structure in an environment which shares many properties with the full dynamical treatment.

Structure can be described both statistically and visually. In Chapter 1, we illustrated the visual aspects of large-scale structure in an elementary treatment of the subject. There are basic aspects of the expanding background model which allow a simplified approach to work on large scales. The metric expansion factor grows, to a good approximation, according to a power law $a(t) \sim t^{2/3}$. The radiation to matter domination transition occurs close to the time of decoupling, after which cosmologically interesting structures are linearly unstable with approximate power law growth $\delta(\mathbf{x}, t) \sim \delta_0(\mathbf{x}) b(t) \sim a(t)$. The structures which arise may be analyzed in various ways. Correlation statistics, pioneered by Peebles (1980) and others, have proven useful in the past and still provide one of the main descriptive tools. These ideas, which set the context for the present study, were addressed in Chapter 2.

The Lagrangian approach to perturbation theory, where material coordinate displacements, rather than spatial fields, provide the perturbative parameters, was described in Chapter 3. The important points of the Zel'dovich approximation for our purposes were its pancaking properties and the common ground it shared with the exact theory.

Simulation methods were described in Chapter 4, as we investigated various initial conditions and the onset of nonlinearity. The power spectrum $P(k)$, the spatial 2-point correlation function in Fourier space, characterized the initial Gaussian density fluctuation field. Small scale filtering (the initial power $P_0(k)$ suppressed for $k \gtrsim k_{nl}$) allowed for a hierarchy of pancaking (with coherence length $\sim 1/k_{nl}$ at each epoch). This was illustrated in one, two, and three dimensions, though we later emphasized three dimensions. Generally, our simulation results fall into two categories: (a) tests for which we have an analytic result and thus with which we can verify the normalization of our simulations and evaluate the dynamic range of our methods, and (b) numerical explorations of the nonlinear or quasi-linear regimes. We showed the agreement of the simulations with the linear theory of local density fluctuations, $\sigma = \langle\delta^2\rangle^{1/2} \approx \sigma_l \sim b(t)$ for $\sigma \ll 1$. Shell crossing in the Zel'dovich approximation caused the fluctuations to stop growing at $\sigma_l \approx 1.5$, a robust result for $P_0 \sim k^n$ spectra with $1 \geq n \geq -3$. We found that $\sigma_l \approx 1$ was a good criterion for placing the simulation in the mildly nonlinear regime, as it was late enough that nonlinearities were important yet early enough that perturbation theory was still applicable. Statistical applications in Chapter 5 emphasized the power $P(k)$ and the reduced bispectrum $Q(\theta)$ (Peebles 1980, Fry & Seldner 1982, Fry 1994b), which is just now being recognized as a practical and potentially important statistic. The bispectrum, sensitive to the anisotropic nature of gravitational instability, is the first correlation function devoted to non-Gaussian features and nonlinear evolution. We were able to verify perturbative effects in power spectrum measurements, such as the first order growth law $P \sim b^2(t)$ and the second order $P \sim b^4(t)$. For a top-hat initial spectrum we were able to directly observe the second and third order mode coupling in three dimensions, though with limited dynamic range. To proceed further, we observed the effect strikingly in one-dimensional simulations, measuring the $P \sim b^{2(n-1)}$ growth for the n loop perturbative corrections. The 3-loop σ^2 calculations of Scoccimarro and Frieman (1996) were verified for $\sigma < 1$ in 3-d simulations. We find that at times perturbative correlation results appear applicable surprisingly far into the nonlinear regime, particularly for the bispectrum. The reduced bispectrum $Q(\theta)$ showed the characteristic enhancements for collinear k_1 and k_2, which would be expected for pancaking. But much more impressively, the second order perturbative result (Fry 1984) applied for scales with $\frac{3}{2}k_1 < k_{nl}$ as might be expected, but it also applied for $k_1 \approx k_{nl}$, particularly if we used the practical technique of inserting the measured $P(k)$ into the perturbative bispectrum expression. The predicted power index n-dependence of Q persisted in the mildly nonlinear regime. On top of this primary effect, the configuration dependence deepened for $n \lesssim -1.5$ and the signal flattened slightly for $n \gtrsim -1.5$. The nonlinear biasing effect (Fry & Gaztanaga 1993) proved to be quite robust, detecting bias in the Q signal during all conditions accessible to our simulations.

REFERENCES

Abell G O 1958 ApJS 3 211

Arnold V I 1992 Catastrophe Theory (Berlin Springer-Verlag)

Arnold V I Shandarin S F & Zel'dovich, Ya. B 1982 Geophys Astrophys Fluid Dynamics 20 111

Bardeen, J M, Bond J R, Kaiser, N, & Szalay A S 1986, ApJ 304, 15

Baumgart, D J & Fry, J N 1991, ApJ, 375, 25

Bernardeau, F 1992 ApJ 392 1

Bernardeau F 1994 AA, 291 697

Bernardeau, F, Singh T P, Banerjee, B & Chitre, S M 1994 MNRAS, 269, 947

Bertschinger, E 1992 in New Insights into the Universe Lecture Notes in Physics, 408 ed V J Martinez M Portilla, & D Saez (Berlin Springer-Verlag), 95

Bertschinger E. 1993 in Relativistic Astrophysics and Particle Cosmology Ann N Y Acad Sci., 688 ed C W Akerlof & M A Srednicki (New York New York Academy of Sciences) 297

Bertschinger E & Dekel A 1989 ApJ 336 L5

Bond, J R. & Szalay A S 1983 ApJ 274 443

Borgani S 1995 Phys Rep., 251, 1

Bouchet, F R Colombi, S., Hivon, E, & Juszkiewicz R. 1995 AA 296 575

Bouchet, F R. Juszkiewicz, R Colombi S & Pellat R. 1992 ApJ 394 L5

Buchert T 1992, MNRAS 254, 729

Coles P 1993 MNRAS, 262, 1065

Coles, P, Melott A & Shandarin, S F 1993 MNRAS 260 765

de Lapparent V Geller M J, & Huchra, J P 1986 ApJ 302, L1

de Vaucouleurs G 1953, AJ, 58 29

Doroshkevich, A. G. 1970, Astrophysics, 6, 320

Doroshkevich, A. G., Ryaben'kii, V. S., & Shandarin, S. F. 1973, Astrophysics, 9, 144

Doroshkevich, A. G., & Zel'dovich, Ya. B. 1963, Astron. Zh., 40, 807

Dressler, A., Dickinson, M., Macchetto, D., Giavalisco, M., & NASA 1994, Space Telescope Science Institute press release STScI PR94-52C

Efstathiou, G. 1990, in Physics of the Early Universe, ed. J. A. Peacock, A. F. Heavens & A. T. Davies (Edinburgh: SUSSP Publications), 361

Einstein, A. 1917, S.-B. Preuss. Akad. Wiss., 142

Friedmann, A. 1922, Z. Phys., 10, 377

Fry, J. N. 1984, ApJ, 279, 499

Fry, J. N. 1994a, ApJ, 421, 21

Fry, J. N. 1994b, Phys. Rev. Lett., 73, 215

Fry, J. N. 1997, Phys. Rep., in preparation

Fry, J. N., & Gaztañaga, E. 1993, ApJ, 413, 447

Fry, J. N., Melott, A. L., & Shandarin, S. F. 1993, ApJ, 412, 504

Fry, J. N., Melott, A. L., & Shandarin, S. F. 1994, MNRAS, 274, 745

Fry, J. N., & Peebles, P. J. E. 1978, ApJ, 221, 19

Fry, J. N., & Scherrer, R. J. 1993, ApJ, 429, 36

Fry, J. N., & Seldner, M. 1982, ApJ, 259, 474

Gamow, G. 1948a, Phys. Rev., 74, 505

Gamow, G. 1948b, Nature, 162, 680

Giovanelli, R., & Haynes, M. P. 1991, Ann. Rev. Astron. Astrophys., 29, 499

Goldston, R. J. & Rutherford, P. H. 1995, Introduction to Plasma Physics (Bristol: Institute of Physics)

Goroff, M. H., Grinstein, B., Rey, S.-J., & Wise, M. B. 1986, ApJ, 311, 6

Gorski, K. M., Hinshaw, G., Banday, A. J., Bennett, C. L., Wright, E. L., Kogut, A., Smoot, G. F., & Lubin, P. 1994, ApJ, 430, L89

Grinstein, B. & Wise, M. B. 1987, ApJ, 320, 448; errata 1989, ApJ, 337, 579

Gurbatov S N, Saichev A I & Shandarin S F 1989 MNRAS 236 385

Harrison E 1970 Phys Rev., D1 2726

Hawking S W & Ellis G F R 1973 The Large Scale Structure of Space-time (Cambridge Cambridge University Press)

Hoyle F 1993 in Observational Cosmology, ed C Chincarini, A Iovino T Maccacaro & D Maccagni (San Francisco Astronomical Society of the Pacific) 693

Hubble E 1929 Proc NAS 15 168

Hubble Space Telescope WFPC Team & NASA 1995 Space Telescope Science Institute press release ST Sci OPO PR95-07

Ichimaru S 1986 An Introduction to Statistical Physics of Charged Particles (Menlo Park. Benjamin/Cummings)

Jain, B & Bertschinger, E 1994 ApJ 431 495

Jeans J 1902 Phil Trans 199A 49

Jeans, J 1928 Astronomy and Cosmogony (Cambridge Cambridge University Press)

Jensen L G., & Szalay, A S 1986 ApJ, 305, L5

Kaiser N 1984 ApJ 284, L9

Kaiser, N & Peacock, J A 1991 ApJ 379 482

Kofman L Bertschinger E Gelb J M Nusser A & Dekel A 1994, ApJ, 420 44

Kofman, L Pogosyan, D Shandarin, S F & Melott A L 1992, ApJ, 393 437

Kolb E W & Turner, M S 1990 The Early Universe (Redwood City, CA Addison Wesley)

Kron R. G 1995 in The Deep Universe, ed B Binggeli & R. Buser (Berlin Springer-Verlag), 233

Lahav O., Naim, A Buta R. J, Corwin, H C de Vaucouleurs C, Dressler A Huchra J P van den Bergh S Raychaudhury S Sodré Jr, L, & Storrie-Lombardi M C 1995 Science 267 859

Lemaître, G 1927 Ann Soc Sci Bruxelles, 47A 49

Lifshitz, E. M 1946 ZhETF 16, 587 J Phys (USSR) 10 116

Lightman A 1991 Ancient Light Our Changing View of the Universe (Cambridge Harvard University Press)

Maddox S J Efstathiou G Sutherland W J. & Loveday J 1990 MNRAS 242 43P

Mather J C. Cheng E S. Cottingham D A., Eplee, R. E Jr Fixsen D J Hewagama, T Isaacman R B Jensen K A. Meyer S S. Noerdlinger P D Read S M Rosen L P Shafer, R. A , Wright, E L. Bennett, C L. Boggess N W. Hauser M G., Kelsall, T , Moseley S H Jr Silverberg, R. F , Smoot G F Weiss R Wilkinson D T 1993 ApJ 420 439

Mattig, W 1958, Astron Nachr 184 109

Melott A. L., Pellman, T F., & Shandarin S F 1994 MNRAS 269 626

Mihalas, D & Binney J 1981, Galactic Astronomy (San Francisco W H Freeman)

Misner C W Thorne K S & Wheeler J A. 1973 Gravitation (San Francisco W H Freeman)

Moutarde, F Alimi J -M., Bouchet, F R., Pellat, R., & Ramani A 1991 ApJ, 382 377

Munshi, D Sahni V , & Starobinsky A A 1994, ApJ 436 517

Naim A , Lahav, O Buta, R. J Corwin Jr , H. G , de Vaucouleurs, G , Dressler, A Huchra J P van den Bergh S Raychaudhury, S., Sodré Jr , L , & Storrie-Lombardi M. C 1995 MNRAS 274, 1107

Pauls, J L & Melott, A L 1995 MNRAS 274 99

Peebles P J E 1971 Physical Cosmology (Princeton Princeton University Press)

Peebles, P J E 1974a, ApJ 189 L51

Peebles, P J E 1974b AA, 32 391

Peebles P J E 1980, The Large-Scale Structure of the Universe (Princeton Princeton Univ Press)

Peebles P J E 1982 ApJ 263 L1

Peebles, P J E. 1993 Principles of Physical Cosmology (Princeton. Princeton Univ Press)

Peebles, P J E & Groth, E J 1975 ApJ 196 1

Peebles P J E , Schramm, D N., Kron R. G , & Turner E. L 1991 Nature 352 769

Peebles P J E & Yu J 1970, ApJ, 162, 815

Penzias A. A & Wilson R. W 1965 ApJ, 142 419

Politzer, H D. & Wise M B 1984 ApJ 285 L1

Press W H, Schechter P 1974 ApJ 187 425

Sahni V & Coles P 1995 Phys Rep 262 1

Sahni V Sathyaprakash B S & Shandarin S F 1994 ApJ 431, 20

Sandage A 1992 Physica Scripta, T43 22

Sandage, A 1993 in Observational Cosmology ed G Chincarini A Iovino, T Maccacaro & D Maccagni (San Francisco Astronomical Society of the Pacific) 3

Scoccimarro R. 1997, ApJ, in press (astro-ph/9812207)

Scoccimarro, R. & Frieman J 1996 ApJS 105 37

Seldner M Siebers, B., Groth E J, & Peebles P J E. 1977 AJ, 82, 248

Shandarin S F Doroshkevich A G & Zel'dovich, Ya. B 1983 Sov Phys Usp. 26, 46

Shandarin S F, & Zel'dovich Ya. B 1989 Rev Mod Phys 61 185

Shane C D & Wirtanen, C A. 1967 Publ Lick Obs 22 Part 1

Shectman, S A Landy S D, Oemler A, Tucker, D L Lin, H., Kirshner, R P, & Schechter, P L 1996 ApJ in press

Smoot G F, Bennett, C L, Kogut A, Wright, E. L., Aymon J Boggess, N W., Cheng, E S, De Amici G, Gulkis S., Hauser M G, Hinshaw G Jackson, P D., Janssen M Kaita, E, Kelsall T, Keegstra, P, Lineweaver C, Loewenstein, K. Lubin, P., Mather J, Meyer S S. Moseley S H Murdock T Rokke L., Silverberg R F Tenorio, L., Weiss, R., Wilkinson, D T 1992 ApJ 396 L1

Thorne K S. 1994 Physics Today, 47 8 61

Truesdell C 1954, The Kinematics of Vorticity (Bloomington Indiana University Press)

Tully, B 1982 ApJ 257, 389

Tully, B 1993, in Observational Cosmology ed G Chincarini A Iovino, T Maccacaro & D Maccagni (San Francisco Astronomical Society of the Pacific) 18

Tully R. B & Fisher J R. 1987 Nearby Galaxies Atlas (Cambridge Cambridge University Press)

Unsold A & Baschek B 1991 The New Cosmos (Berlin Springer-Verlag)

Weinberg, S 1972 Gravitation and Cosmology Principles and Applications of the General Theory of Relativity (New York Wiley)

Weinberg S 1977, The First Three Minutes (New York Basic Books)

Zel'dovich, Ya B 1965 Adv Astron Astrophys 3 241

Zel'dovich Ya B 1970 AA, 5 84

Zel'dovich Ya. B 1972 MNRAS 160 1P

Zel'dovich Ya. B 1983 Proc NAS, 80 2410

Zel'dovich Ya. B 1992, My Universe Selected Reviews ed Ya B Zel'dovich & M V Sazhin (Chur Harwood Academic Publishers)

Zel'dovich Ya B 1993 Selected Works of Yakov Borisovich Zel'dovich, Volume II Particles Nuclei and the Universe ed J P Ostriker (Princeton Princeton University Press)

Zel'dovich Ya. B & Novikov I D 1983 The Structure and Evolution of the Universe ed G Steigman (Chicago University of Chicago Press)

Zwicky, F , Herzog, E Wild, P Karpowicz, M & Kowal C T 1961–68, Catalogue of Galaxies and Clusters of Galaxies, in 6 vols (Pasadena. California Institute of Technology)

BIOGRAPHICAL SKETCH

The author received a B.S. in physics from M.I.T. in Cambridge, Massachusetts. His interests include physics, computers, bass and drums, cats, and people.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James N. Fry

James N. Fry, Chair
Professor of Physics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Paul Avery

Paul R. Avery
Professor of Physics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James R. Ipser
Professor of Physics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Pierre Sikivie
Professor of Physics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Zoran R. Pop-Stojanovic
Professor of Mathematics

This dissertation was submitted to the Graduate Faculty of the Department of Physics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1997

Dean, Graduate School